Dottorato di ricerca in
Metodologia Statistica per la Ricerca Scientifica
XIX ciclo

Università di Bologna

Invariants estimation in nonlinear time series

Laura Sardonini

Dipartimento di Scienze Statistiche "P. Fortunati"
Marzo 2007

Dottorato di ricerca in
Metodologia Statistica per la Ricerca Scientifica
XIX ciclo

Università di Bologna

Invariants estimation in nonlinear time series

Laura Sardonini

coordinatore

prof. Daniela Cocchi

tutor

prof. Rodolfo Rosa

Settore Disciplinare
SECS-S/02

Dipartimento di Scienze Statistiche "P. Fortunati"
Marzo 2007

Preface

With this work we would like to investigate some particular phenomena appearing messy and without control. In fact a "*prima facie*" behavior of phenomena looks like without rules and principles. The aim of the research is to understand phenomena which it is possible to meet in nature. The starting point of the research in this sense can be found in the investigation the falling stars and the eclipses. Isaac Newton was the first one to define laws governing phenomena for planet motion and he found out "deterministic" laws. A law is called deterministic if future and past are made unique by present state. The main discover is that perfect deterministic law can create a chaotic and unpredictable motion. This kind of motion is called "*deterministic chaos*" and it could represented the third scientific revolution after theory of relativity and quantum mechanics.

One of the main characteristic of chaotic deterministic systems is the absence of making forecast in a long period, furthermore the propriety depends on the system complexity. The found solutions are so complicated to look like random. Therefore we can assert: "*chaos*" is not so far away from "*determinism*". Concepts of "*chaos*" and "*determinism*" have been developed in a short period of time because of multiple interests in several fields of science (examples of such systems include the atmosphere, the solar system, the plate tectonics, the turbulent fluids, economics, and the population growth).

The thesis is organized as follows.

In the introduction we will discuss some general aspects of the time series analysis related to chaotic time series, theory of dynamical system, ergodic theory and some invariant aspects of the time series. Chapter 2 will give a review of some widely used methods for estimation of the fractal dimension of a chaotic attractor from a corresponding time series.

Chapter 3 will describe the open question about the choice of values for the scaling region. Some methods to carry out the region will also be discussed. The main purpose of this section is how the identification problem is related to the scaling region. Chapter 4 will introduce the U-Statistics theory and some basic results. Chapter 5 will present some theoretical results using the U-Statistic theory to Gaussian Correlation Integral and the Takens estimator. Chapter 6 will focus on a problem of estimating the variance of the sample gaussian correlation integral using the U-Statistics. The results are supported by the literature. Then, aim of the chapter is to find the correlation dimension distribution.

Analysis have been obtained using the software Matlab.

Acknowledgements

This thesis is the result of my three-years period of research at the Statistics Department of the University of Bologna.

It is my great pleasure to record thanks to people who helped in many ways to contribute to this thesis.

I would like to thank Prof. Svetlana Borovkova, who welcome me at the Mathematics Department of Technology University of Delft. She helped and encouraged me in my PhD program.

I would like to thank to my supervisor Prof. Rodolfo Rosa, for introducing me to many ideas on the area of chaotic dynamical systems and chaotic time series and also for his critical guidance, support and inspiration and for never letting me down during the hard moments of doing research.

Many thanks go to Dr. Cees Diks from Amsterdam University for his advices and suggestions. Thanks to Dr. Bart Hoekstra from Leiden University Medical Center for his interest towards my work.

My period of study at Mathematics Department of Delft was crucial for my research, so I would like to thank all the people made it possible. A special thank to Federico Lucidi, Xu and Daria Spivakovskaya. I am grateful to my colleagues in Bologna. My special thanks to (alphabetic order) Valentina Adorno, Marta Disegna and Mariagiulia Matteucci for the fun time we had together. It is impossible to forget the elder colleagues Caterina Liberati and Giulia Roli for advice and suggestion. Thanks girls!

Many thanks to all people I had close to me in these years. A very special thank go to my best friends Giulia and Marco for their patience and support. Thanks to my family for trusting me. Thanks to Marino for holding on. A special thank to a new soul, met under the rain, for his smiles and his presence in my life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The meaning of "chaos" is connected to the idea that small causes can sometimes have large effects; a clear example is given in this popular sentence "*for want of a nail ... a kingdom was lost*" as ancient people said.

The presence in nature[1] of phenomena appearing without rules gave the idea of finding out some methods to understand the behaviors. In this section, we will show what happens for some phenomena by using the physical field, in fact everyone experienced once in life some physical laws which produce a periodic and regular evolutions. In the reality there are phenomena with unseen rules. The answer to this problem can be found in the use of the phase space[2] which provides answers to a large set of questions for example how one can extract useful physical information from observations of chaos. Formally, chaos theory is defined as the study of complex nonlinear dynamic systems, the word nonlinear implies recursion and higher mathematical algorithms and dynamic implies non-constancy and non-periodicity. Thus, chaos theory is the study of complex systems based on mathematical concept of recursion, whether in the form of a recursive process or a set of differential equations modelling a physical system.

The most commonly held misconception is that chaos theory is about disorder. Nothing could be further from the truth. It does not disprove determinism or dictate that ordered systems are impossible; it does not invalidate experimental evidence or claim that modelling complex systems is useless. The "chaos" in chaos theory is order not simply order, but the very *essence* of order.

It is true that chaos theory dictates that minor changes can cause huge fluctuations. But one of the central concepts of chaos theory is that while it is impossible to exactly

[1] We refer to Physics, Biology, Sociology and Economics.

[2] Phase space: particular space where it is possible to observe the evolution see next paragraph.

predict the state of a system, it is generally quite possible, even easy, to model the overall behavior of a system. Thus, chaos theory lays emphasis not on the disorder of the system or the inherent unpredictability of a system, but on the inherent order in the system (the universal behavior of similar systems).

Thus, it is incorrect to say that chaos theory is about disorder. To take an example, consider Lorenz's Attractor. The Lorenz Attractor is based on three differential equations, three constants and three initial conditions. The attractor represents the behavior of gas at any given time and its condition depends on a previous time. If the initial conditions are changed by even a tiny amount, say as tiny as the inverse of Avogadro's number (a heinously small number with an order of $1 \cdot 10^{-24}$), checking the attractor at a later time will yield numbers totally different. This is because small differences will propagate themselves recursively until numbers are entirely dissimilar to the original system with the original initial conditions. However, the plot of the attractor will look very much the same. Both systems will have totally different values at any given time and the plot of the attractor (the overall behavior of the system) will be the same. This system will be consider in the following.

The basic idea of chaos theory is that complex nonlinear systems are inherently unpredictable. At the same time, chaos theory also insures that the way to express such an unpredictable system does not often lie in exact equations, but in representations of the system behavior in terms of strange attractors or in fractals. Many people think that chaos theory is about unpredictability instead of chaos theory is at the same time about predictability even in the most unstable systems.

1.1 Definition of chaos

A more precise definition of chaos is given in this section, by using working definitions. In fact we will look at certain nonlinear dynamical systems that, under certain conditions, show "chaotic" behavior. Among the characteristics of chaotic systems, described below, the sensitivity to initial conditions is the more evident. A result of the sensitivity condition is that the behavior of system appears to be random, even though the system is deterministic in the sense that it is well defined and contains no random parameters.

The working definitions used focus on the general characteristics of chaos motion so it is possible to describe the setting in which we will work. The main characteristics of chaos are summarized as follows:

▷ *Determinism*

Theoretically, the chaos motion must be generated by one or more deterministic equations that do not contain any random factors. The system states of past, present and future are controlled by deterministic rules. Many scientists have found stochastic-like behavior in completely deterministic system so they have been attracted to probe into the secrets of the system.

▷ *Nonlinearity*

Nonlinearity is a necessary, but not sufficient condition for the appearance of chaos. Chaos motions must come from a nonlinear system but nonlinearity does not necessarily imply chaos.

▷ *Sensitive Dependence on Initial Conditions (SIDC)*

Generally, the evolution of a system depends on its initial state. When our interest focuses on the whole dissipative attractor, the evolution seems not sensitively dependent on its initial states because all trajectories fall onto the attractor. But if our interest focuses on the inner structure of the strange attractor, we can induce that the path of trajectories diverge and converge exponentially. For a chaotic system, this property must be valid for nearly almost all possible initial states. Under this constraint, a possible geometrical explanation for chaotic structures is stretching and folding. The chaotic trajectories move within the finite phase space[3] forever.

▷ *Aperiodicity*

Chaotic motion is a new topological type of motion that is very different from fixed point, limit cycle and limit torus. Its orbits are non-periodic. This means that a chaotic orbit can never join another one or repeat its history. But not all non-periodic orbits are chaotic orbits. Almost periodic motions and quasi-periodic motions are aperiodic but not chaotic.

▷ *Stability with Some Tension and Boundness*

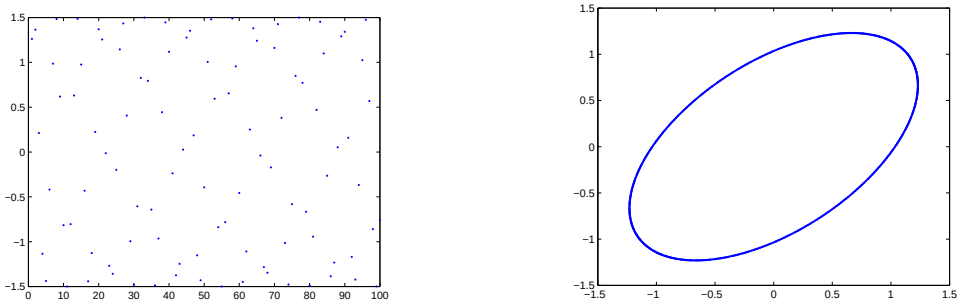
There are different approaches towards the requirements of the chaos stability. In pure mathematics a definite chaos without consideration of stability is convenient. In Physics the stability is so important that it is better to include stable constraints in defining chaos motions. In fact, scientists tend to understand the chaos motion physically. In their eyes, chaos is bound and involves a kind of loose stability. The chaos motion on the strange attractor is locally unstable but globally stable.

[3] see pag. 13

Lastly, the classic chaos can be defined as a recurrent, random-like and aperiodic behavior generated from deterministic nonlinear equations with sensitive dependence on initial conditions of the system. The statistical interest is justified by the random-like behaviors of phenomena. For this reason there will be the possibility to estimate characteristics of chaos and its accuracy levels.

1.2 Examples

We are interested to show some simple and famous examples and to look at what above proprieties mean. First of all we can take into account a mathematical function given by the sine wave $f(t) = Asin(t)$. See in the one dimension Figure 1.1(a) (the $f(t)$ space) this oscillates between $\pm A$.



(a) (b)

Figure 1.1. Sine wave in one and two dimension space for $A = \pm 1.5$

Two points in this line which are close in the sense of Euclidean or other distance may have quite different values of $\dot{f}(t)$. So two close points in one dimension may be moving in opposite directions along the single spatial axis chosen for viewing the dynamics. Seen in two-dimensional space $[f(t), f(t+T\tau_s)]$ the ambiguity of the velocity of the points is resolved, and the sine wave is seen to be motion on a figure topologically equivalent to a circle. It is generically an ellipse whose depends on the value of T see Figure 1.1(b). The overlap of orbit points due to projection onto the one-dimensional axis is undone by the creation of the two-dimensional space.

Another example of chaos is given by the logistic model popularized in a seminal 1976 paper by the biologist Robert May. The logistic model was originally introduced

as a demographic model by Pierre François Verhulst. Later it was applied on surplus production of the population biomass of species in the presence of limiting factors such as food supply or disease, and so two causal effects:

▷ reproduction means the population will increase at a rate proportional to the current population

▷ starvation means the population will decrease at a rate proportional to the value obtained by taking the theoretical "carrying capacity" of the environment less the current population.

For example, the spread of disease can be modelled by a relatively simple function called the Logistic Equation. The spread of disease depends on many things: the number of people initially infected, the number of carriers (people who have the virus but do not show symptoms), the number of people not infected, whether or not the disease is curable, etc..It is possible to write this model mathematically so we have

$$x_{t+1} = rx_t(1 - x_t),\ t = 1, 2, \ldots \tag{1.1}$$

where:

▷ x_t is a number between zero and one, and represents the population at year t, and hence x_0 represents the initial population (at year 0)

▷ r is a positive number, and represents a combined rate for reproduction and starvation.

The relative simplicity of the logistic map makes it an excellent point of entry into a consideration of the concept of chaos. When we plot points of known system, the behavior looks like an uniform distribution. It is possible to see what we have said in the Figure 1.2.

A rough description of chaos is that chaotic systems exhibit a great sensitivity to initial conditions – a property of the logistic map for most values of r between about 3.57 and 4 (as noted above). A common source of such sensitivity to initial conditions is that the map represents a repeated folding and stretching of the space on which it is defined. In the case of the logistic map (1.1), the quadratic difference equation describing it may be thought of as stretching and folding operation on the interval $[0, 1]$. To see the behavior we use the scatter-plot of the logistic map which is a special case of the phase space reconstruction using the embedding theorem (Th.1 14). Two and three

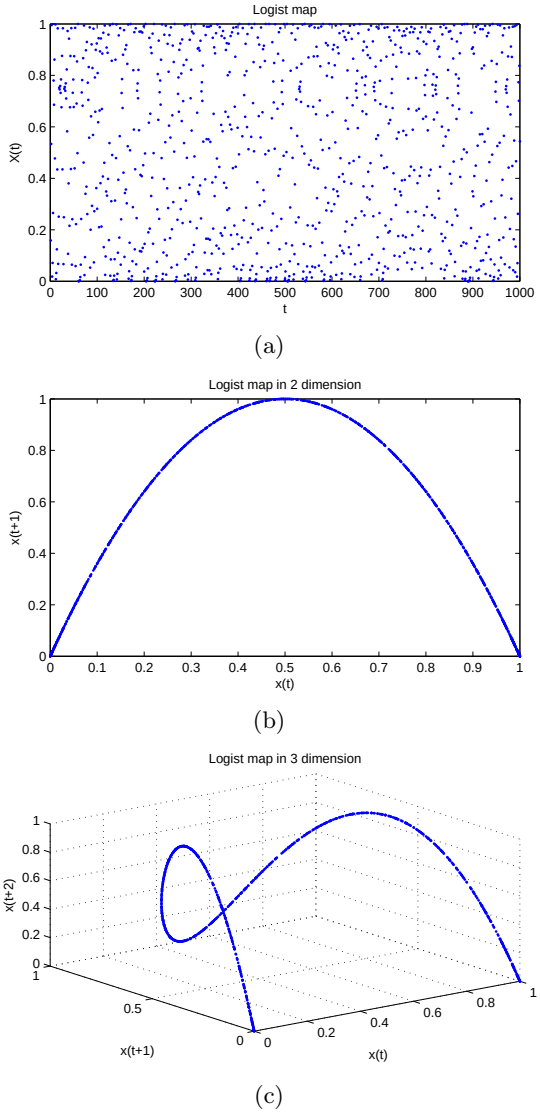


Figure 1.2. Logistic map in different space: (a) Logistic map in the original space, (b) Scatter-plot of Logistic map (2 dimension), (c) Logistic map in 3 dimension

dimensional phase diagrams show the stretching and folding structure of the logistic map.

The Figure 1.2 illustrates the stretching and folding over a sequence of iterates of the map. Figure in the middle gives a two-dimensional phase diagram of the logistic map for $r = 4$, and clearly shows the quadratic curve of the difference equation (1.1). However, we can embed the same sequence in a three-dimensional phase space, in order to investigate the deeper structure of the map. The last figure demonstrates this,

showing how initially nearby points begin to diverge, particularly in those regions of x_t corresponding to the steeper sections of the plot.

This stretching and folding does not just produce a gradual divergence of the sequences of iterates, but an exponential divergence (see Lyapunov exponents), evidenced also by the complexity and unpredictability of the chaotic logistic map. In fact, exponential divergence of sequence of iterates explains the connection between chaos and unpredictability: a small error in the supposed initial state of the system will tend to correspond to a large error later in its evolution. Hence, predictions about future states become progressively (indeed, exponentially) worse when there are even very small errors in our knowledge of the initial state.

It is often possible, however, to make precise and accurate statements about the likelihood of a future state in a chaotic system. If a (possibly chaotic) dynamical system has an attractor, then there exists a probability measure that gives the long-run proportion of time spent by the system in the various regions of the attractor. In the case of the logistic map with parameter $r = 4$ and an initial state in $(0, 1)$, the attractor is also in the interval $(0, 1)$ and the probability measure corresponds to the beta distribution with parameters $a = 0.5$ and $b = 0.5$. Unpredictability is not randomness, but in some circumstances looks very much like it. Hence, and fortunately, even if we know very little about the initial state of the logistic map (or some other chaotic system), we can still say something about the distribution of states a long time into the future, and use this knowledge to inform decisions based on the state of the system.
In the Logistic model example we use the scatter-plot as a new space to see the evolution of the system. In nonlinear time series and in particular in chaotic ones the analysis is made using the concept of *phase space*. It is used in mathematics and physics where *phase space* is the space in which all possible states of a system are represented, with each possible state of the system corresponding to one unique point in the phase space.

In phase space, every degree of freedom or parameter (number of variables needed to explain the system) of the system is represented as an axis of a multidimensional space. For every possible state of the system, or allowed combination of values of the system's parameters, a point is plotted in the multidimensional space. Often this succession of plotted points is analogous to the system's state evolving over time.

For simple systems, such as a single particle moving in one dimension for example, there may be as few as two degrees of freedom, (typically, position and velocity), and a sketch of the phase portrait may give qualitative information about the dynamics of system.

The theoretical justification of the phase space is given by the Takens[4] theorem also called delay embedding theorem. The theorem establishes conditions under which a chaotic dynamical system can be reconstructed from a sequence of observations of the state of a dynamical system. The reconstruction preserves the properties of the dynamical system that do not change under smooth coordinate changes, but it does not preserve the geometric shape of structures in phase space.

Delay embedding theorem is simpler to state for discrete time dynamical systems. The main idea of the theorem is to use a multidimensional space instead of the original one to observe the time series. The main idea is to use the scalar observation and to go from this space to multivariate phase space thanks the embedding theorem.

Let (X, T) be a dynamical system with finite-dimensional state space X and bound positive orbits $\{T^n x\}_{n\geq 0}$. Let $f : X \to \mathbf{R}$ be a real-out function. Usually $f(x)$ represents a real-valued measurement made on a point $x \in X$. If $(x_0, Tx_0, T^2x_0, \ldots)$ is an orbit of the dynamical system with initial state x_0, then the corresponding time series is obtained by applying to each point of the orbit: $(f(x_0), f(Tx_0), f(T^2x_0), \ldots)$.

Let both T and f be continuously differentiable. Define the vector-valued reconstruction map $y : X \to \mathbf{R}^m$ by

$$y_i = (f(x_i), f(Tx_i), \ldots, f(T^{m-1}x_i)) \in \mathbf{R}^m \tag{1.2}$$

Theorem 1 (Takens-Mañé). *In the cartesian product of the space of C^1-mapping on X and the space of C^1-functions from X to $\boldsymbol{R}$ there exists an open and dense subset U, such that if $(T, f) \in U$, then the reconstruction map y_i, defined in 1.2, is an embedding, whenever $m > 2 \cdot dim(X)$.*

These y replace the scalar data measurements with data vectors in an Euclidean m-dimensional space in which the invariant aspects of the sequence of points x are captured with no loss of information about the proprieties of the original system.

The new space is related to the original space of the x by smooth, differentiable transformations. The smoothness is essential in allowing the demonstration that all the invariants of the motion as seen in the reconstructed time delay space with data y which are the same as if they were evaluated in the original space. This means we can work in the reconstructed time delay space and learn essentially as much as we could about

[4]Next to Takens even Mañé showed same results.

the system at the source of our observations as if we were able to make our calculation directly in the *true* space.

1.3 Chaotic dynamical system and ergodicity

In the above section we will show some characteristics and definitions related to chaos systems. Here we want to discuss in a more formal way some aspects of chaotic dynamical systems. We will focus on systems with discrete time.

A discrete time dynamical system (X, T) is a pair consisting of the state space X, the set of all possible values, and the time evolution map $T : X \to X$, the law according to which a state evolves to other states at later times. For an initial state $x_0 \in X$, the iterations of T give rise to a trajectory, or an orbit $\{T^m x_0\}_{n \in N}$ (or, if T is invertible, $n \in Z$)

A dynamical system is related to a time series by means of the read-out function, or the observable function $f : X \to R$ which assigns to each possible state in X the recorded value when the system is in that state.

The absence of unique definition of chaos is showed in formal way by the lack of unique definition of attractor on dynamical systems. An attractor is the limiting set where the experimental orbits $\{T^m x\}_{n \in N}$ accumulate for large n. A more precise definition is that of an attracting set. The set A is called attracting set with fundamental neighborhood U, if it satisfies the following properties:

1. *Attractivity*: for every open set $V : A \subset V$ we have $\{T^m x : x \subset U\} \subset V$ for all sufficiently large n;

2. *Invariance*: for all $x \in A$ and all n we have $T^n x \in A$

The notion of invariant measures is associated with a dynamical system. A finite measure μ on the Borel σ-field F of X is called T-*invariant* if, for any set $B \in F$, $\mu(T^{-1}B) = \mu(B)$. Since μ is finite, we can assume without loss of generality that μ is a probability measure, i.e. that $\mu(X) = 1$. To indicate the link between the dynamical system and the corresponding invariant measure we shall sometimes write (X, T, μ) for a dynamical system.

A transformation T (if it is a homeomorphism) always has at least one invariant measure associated with it (if the state space is compact); in fact it can have more. Typically, there are many invariant measures on an attractor. Of particular interest for us are the so-called ergodic measure.

A *T-invariant measure* μ is called ergodic if all T-invariant sets in X (i.e. all $A \in F$ for which $T^{-1}(A) = A$) have measure $\mu(A)$ either 0 or 1.

An attractor and an invariant measure provide us a global description of the asymptotic behavior of a dynamical system. The dynamic on the attractor itself does not need to be simple. For some dynamical systems the evolution on the neighborhood of the attractor may depend sensitively on initial conditions, i.e. the trajectories starting in nearby initial points diverge from each other at an exponential rate and after some time can be found in totally different parts of the attractor. This property is what we mean with the sensitive dependence on initial conditions.

Recall that a dynamical system together with an invariant measure is called *ergodic* if all T-invariant sets have the measure μ of 0 or 1. An ergodic dynamical system may possess some mixing property with respect to the invariant measure. Many different types of mixing conditions can be considered, such as weak or strong mixing condition, weak Bernoulli, etc. The essential meaning of mixing is that the future evolution becomes almost independent of the past, as time goes by. In particular the convergence of mixing coefficients indicates that the fact that, when taken far apart, the past and the future of the process become almost independent, and the sequence of data is expected to exhibit an asymptotic behavior which is close to that of an independent sequence.

1.4 Invariants

The presence of a smooth phase space gives us the possibility to study some invariant quantities in a easier way. So in this new space we can study some proprieties known as *fractal structural* and its direct consequence the *fractal dimension*, *Lyapunov exponents* which are defined in the state space. In the next we will show a short presentation about *fractal dimension* and *Lyapunov exponents*.

1.4.1 Fractal dimension

The term *fractal* has no precise meaning in the scientific literature, although most people even somewhat familiar with the expression. Probably they associate it with complicated irregular subsets of the plane. *Fractal dimension* can be regarded as one measure of the complexity or irregularity of these striking images; it can also be interpreted as the degree to which a set fills the Euclidean space in which it is embedded. Notions of fractal dimension have been connected to probability distributions and dynamical

systems and found use as tools in helping scientists and data analysis distinguishes deterministic processes from stochastic phenomena.
The term *fractal* originated with Benoit Mandelbrot who noted that many objects occurring in nature are highly irregular in structure (e.g. coastlines, snowflakes, crystal) yet in possession of a certain degree of self-similarity. By the latter we mean that the object, viewed at increasing levels of manifestation, appears to be a union of many ever smaller "copies" of itself. Irregular self-similar sets provide the simplest, most basic examples of fractals.

Example The standard mathematical example of an irregular self-similar object is the ternary Cantor set in $[0, 1]$. It is constructed by a very simple algorithm. First remove the open middle third $(\frac{1}{3}, \frac{2}{3})$ from $[0, 1]$, then remove the open middle thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ from the resulting two line segments. Continue this process indefinitely, at each stage removing the open middle thirds from any remaining line segments from the previous stage. The set of points that remains in the limit is called the Cantor set, and it is frequently referred to as *Cantor dust*, to emphasize its relative scarcity in $[0, 1]$. It is a closed set (being the intersection of a sequence of closed sets) and with Lebesgue measure equal to zero (this seen by observing that the lengths of the removed open intervals in the construction sum to 1). We also note that it is self-similar, since it can be split as the disjoint union of two smaller copies of itself. The geometric irregularity of the Cantor set is perhaps less obvious, since we cannot actually see the set. However, topologically, it can be described as abundant yet "full of holes". Note that a point belongs to the Cantor set if and only if it has a base 3 expansion consisting entirely of 0's and 2's (removing the middle third intervals eliminates the 1's). Hence, in terms of cardinality, the Cantor set is uncountable and every point of the set is a limit point of a sequence of other points of the set. This indicates abundance of a sort, on both a global and local level. Yet the construction also shows that any two points of the set are separated by a missing open, interval, so the set is full of gaps. We can see an example in the Figure 1.3

The example uses a line segment but it is possible to expand the method for a square or cube, etc. . . and recursively removing portions of the set remaining at each stage is often called a *Cantor-like construction*. Here we can consider the unit square Figure 1.4 . We can divide the unit square into a 3×3 grid of nice smaller squares, each of side length 1/3. We then remove the four corner squares and the middle square, producing another set. We can do the same $n-times$ removing the corner and the middle squares. It can easily see that the set is rapidly vanishing while becoming more irregular and

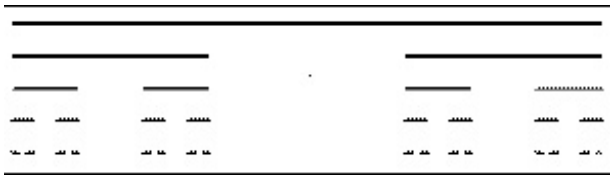


Figure 1.3. Cantor set in $\mathbb{R}$

fractured. The self-similarity is also evident. In this case the set is a union of four non-overlapping copies of itself, each copy reduced by the factor one-third. Every set which follow this construction structure is called *Cantor-like set*.

Here we want to give an idea about the fractal dimension, there a lot of definitions. One of the definitions is the *Box-dimension* and this is the simplest one.

Definition 1 (Box Dimension or Box-Counting Definition). *Let E be a nonempty bounded subset of R^N, and $N_\epsilon(E)$ is the minimum number of closed balls of diameter ϵ required to cover E. Then the lower and the upper box dimension of E are defined to be*

$$\begin{aligned} \Delta^-(E) &= \liminf_{\epsilon \to 0} \frac{\log N_\epsilon(E)}{\log 1/\epsilon} \\ \Delta^+(E) &= \limsup_{\epsilon \to 0} \frac{\log N_\epsilon(E)}{\log 1/\epsilon} \end{aligned} \tag{1.3}$$

If $\Delta^-(E) = \Delta^+(E)$ then we call this common value the box dimension of E and denote it by $\Delta(E)$.

The box dimension is defined via the minimum number of closed balls required to cover the attractor. Using this definition we can compute the box dimension for the Cantor set C in Figure 1.3. The middle-third construction of C suggests to cover $[0, 1]$ with the family of grids consisting of the triadic intervals $[j3^{-n}, (j + 1)3^{-n}]$. We then easily count $N_n(C) = 2^n$ and $\Delta(C) = \log 2^n / \log 3^n = \log 2 / \log 3 = 0.6309$.

Other dimensions studied in the literature are: the Haussdorf dimension, the information dimension, the correlation dimension. For a good review on the question of dimension see Cutler [1993].

For purpose of statistical estimation, the correlation dimension is most appropriate because it characterizes the invariant measure on the attractor, and it is relatively easy

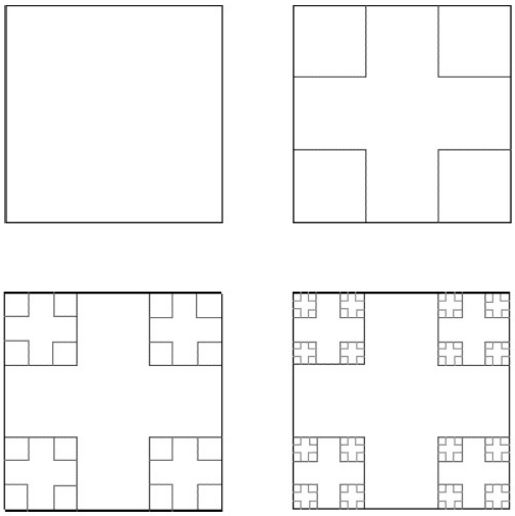


Figure 1.4. Cantor set in $\mathbb{R}^2$

to estimate. So we decide to focus on the correlation dimension.

1.4.2 Lyapunov exponents

A quantitative measure of the sensitive dependence on the initial conditions is the Lyapunov exponent λ. It is the averaged rate of divergence (or convergence) of two neighboring trajectories. Quantitatively, two trajectories in phase space with initial separation $\delta\mathbf{Z}_0$ diverge following the next relation:

$$|\delta\mathbf{Z}(t)| \approx e^{\lambda t}|\delta\mathbf{Z}_0|$$

The rate of separation can change for different orientations of initial separation vector. Thus, there is a whole spectrum of Lyapunov exponents the number of them is equal to the number of dimensions of the phase space. It is common to just refer to the largest one, because it determines the predictability of a dynamical system.

For a dynamical system with evolution equation f^t in a $m-$dimensional phase space, the spectrum of Lyapunov exponents

$$\{\lambda_1, \lambda_2, \cdots, \lambda_m\}$$

in general, depends on the starting point x_0. This behavior can be shown in the

Figure 1.5, taken two close points after a period of time t their distance is not constant but it grows exponentially.

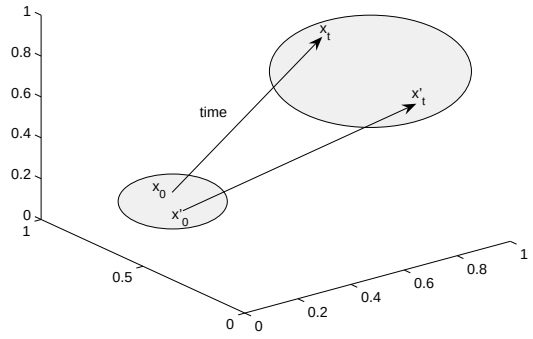


Figure 1.5. Example of divergence for two close points during the trajectory.

The Lyapunov exponents describe the behavior of vectors in the tangent space of the phase space and they are defined from the Jacobian matrix

$$J^t(x_0) = \left.\frac{df^t(x)}{dx}\right|_{x_0}$$

The J^t matrix describes how a small change at the point x_0 propagates to the final point $f^t(x_0)$. The limit

$$\lim_{t\to\infty}(J^t \cdot (J^t)^T)^{1/2t}$$

defines a matrix $L(x_0)$ (the conditions for the existence of the limit are given by the Oseldec theorem). If $\Lambda_i(x_0)$ are the eigenvalues of $L(x_0)$, then the Lyapunov exponents Λ_i are defined by

$$\lambda_i(x_0) = \log \Lambda_i(x_0)$$

The set of Lyapunov exponents will be the same for almost all starting points of an ergodic component of the dynamical system. If the system is conservative (i.e. there is no dissipation), a volume element of the phase space will stay the same along a trajectory. Thus the sum of all Lyapunov exponents must be zero. If the system is

dissipative, the sum of Lyapunov exponents is negative.

If the system is a flow, one exponent is always zero the Lyapunov exponent corresponding to the eigenvalue of L with an eigenvector in the direction of the flow.

Generally the calculation of Lyapunov exponents, as defined above, cannot be carried out analytically, and in most cases one must resort to numerical techniques. The commonly used numerical procedures estimates the L matrix based on averaging several finite time approximations of the limit defining L. For the calculation of Lyapunov exponents from limited experimental data, various methods have been proposed.

Whereas the (global) Lyapunov exponent gives a measure for the total predictability of a system, it is sometimes interesting to estimate the local predictability around a point x_0 in phase space. This may be done through the eigenvalues of the Jacobian matrix $J^0(x_0)$. These eigenvalues are also called local Lyapunov exponents. The eigenvectors of the Jacobian matrix point in the direction of the stable and unstable manifolds.

Chapter 2

Review of correlation dimension estimators and methods

In this chapter we want to investigate the most popular estimators for the correlation integral and the correlation dimension. The analysis of one of these quantities is related with the other one because correlation dimension is a function on the correlation integral. For this reason we will give both estimators definitions in this chapter. For each estimators it is possible to associate a particular estimating method. The differences between these methods will also be the object of this chapter. Among the various notions of the dimension, the correlation dimension receives the most attention in the literature, mainly because of its relatively easy computation.

2.1 Introduction

Let $\mathcal{F}$ be an ergodic probability measure for a dynamical system $\{X_t, t \in Z_+\}$ that lives on attractor $A \subseteq R^m$. The spatial correlation related to the dynamical system can be describe in terms of the correlation integral $C_{\mathcal{F}}(\epsilon)$ which is the probability distribution function of the inter-point distance $|X-Y|$ where X and Y are independent and identically distributed with $\mathcal{F}$ as their common probability measure. This can seen from the following identities:

$$P(|X-Y| \leq \epsilon) = E(P(|X-Y| \leq \epsilon|Y)) = C_{\mathcal{F}}(\epsilon) \tag{2.1}$$

Note that it is common in the dynamical literature to adopt the maximum norm in the case of vector variables because the computation is more convenient because the correlation dimension does not depend on which vector norm is adopted. Correlation integral represents the probability that the distance between the pairs of points (x_i, x_j) is less than ϵ.

As an example, let the natural measure be $U(0.1)$, the uniform distribution on $[0, 1]$. Then it can be found the value of $C(r)$. First note that the joint distribution of X and X' is the uniform distribution on the square $[0, 1] \times [0, 1]$. On the axis it is possible to find the single event, so if events have distance equals to zero, all possible combinations lie over the diagonal. Otherwise all possible combinations of points with maximum distances less than r lie in the black area in Figure 2.1. The probability that $|X - X'| \leq r$, (black area) is equals one minus the area of the two white triangle corners. So, $C(r) = P(|X - X'| \leq r) = 1 - (1 - r)^2 = 2r - r^2$

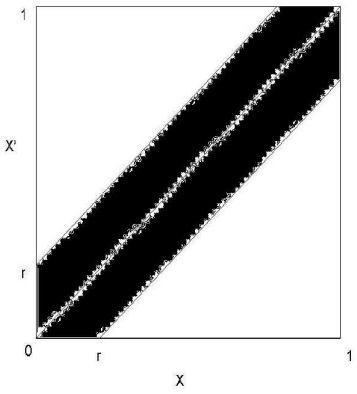


Figure 2.1. Probability that $|X - X'| \leq r$ is equal the inner (black) area two dashed lines

In general the computation of invariants is made considering $\{\mathbf{x_i}\}$ as an embedding time series of a particular system with an appropriate embedding dimension m and time delay τ. So the correlation integral is defined in terms of the distribution $f_m(\mathbf{x})$ of delay vectors as

$$C_{\mathcal{F}}(\epsilon) = \int\int \Theta(\epsilon - |\mathbf{x} - \mathbf{y}|) f_m(\mathbf{x}) f_m(\mathbf{y}) d\mathbf{x} d\mathbf{y} \tag{2.2}$$

The correlation dimension α is defined as the exponent appearing in the scaling law:

$$C_{\mathcal{F}}(\epsilon) \sim \epsilon^{\alpha} \tag{2.3}$$

To be more specific, the above scaling law requires that

$$\lim_{\epsilon \to 0} \frac{\log C_{\mathcal{F}}(\epsilon)}{\log \epsilon} = \alpha \tag{2.4}$$

A log-log plot of $C_{\mathcal{F}}(\epsilon)$ is approximately linear for small ϵ and the asymptotic slope being defined as the correlation dimension. Note that α can be any real number between 0 and m. It is revealing to assign the scaling law in the product form:

$$C_{\mathcal{F}}(\epsilon) = \Phi(\epsilon)\epsilon^{\alpha} \tag{2.5}$$

where the pre-factor $\Phi(\epsilon)$ satisfies the requirement that

$$\lim_{\epsilon \longrightarrow 0} \frac{\log \Phi(\epsilon)}{\log \epsilon} = 0 \tag{2.6}$$

The scaling law is said to be exact if the pre-factor Φ is constant over the range $0 \leq \epsilon \leq \epsilon_0$ for some positive ϵ_0. Ordinarily, the scaling law is not exact. The condition needed is that the pre-factor should be asymptotically constant and it may oscillate or even become unbounded as ϵ goes to zero. The phenomenon of an asymptotically non-constant pre-factor is referred to as *lacunarity*.

The term lacunarity was first introduced by Maldelbrot as a measure of the texture of a fractal set. Lacunarity represents a counterpart to the fractal dimension and describes the degree of gappiness of a fractal. It is strongly related to the size distribution of the holes on the fractal and to its deviation from translation invariance; roughly speaking, a fractal is very lacunar if its holes tend to be large, in the sense that they include large region of space. If a fractal has large gaps or holes, it has high lacunarity; on the other hand if a fractal is almost translationally invariant, it has low lacunarity. Different fractals can be constructed that have the same dimension but that look widely different because they have different lacunarity. For this reason an oscillating pre-factor complicates the estimation of the correlation dimension. A graphic explanation of

lacunarity is given by Figure 2.1

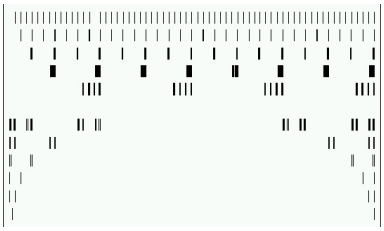


Figure 2.2. Example of lacunarity in a set: the holes represent the degree of gappiness

When lacunarity is absent it is possible to define the background as the ideal case so it is simple to have some estimators without consider this factor. One of these is the Takens estimator where $\Phi(\epsilon)$ is assumed constant. Therefore, an efficient and unbiased estimate for α exists.

Formula (2.1) and (2.4) are the definitions for quantities which represent our interest; in the next part we will see the most popular estimators.

2.2 Grassberger-Procaccia Estimator

The correlation integral (2.1) was initially introduced by Grassberger-Procaccia as an efficient estimator to Box-counting dimension $\Delta_\epsilon(E)$ see page 18. It is easy and comparatively quick to compute, and allows the experimentalist to probe the attractor at much smaller distances than does box-counting. Consequently, the correlation integral method (usually in conjunction with the time-delay embedding) remains one of the most popular of the dimension estimation techniques currently in use.
If $\mathbf{x_1}, \mathbf{x_2}, \ldots$ is a sequence of random vectors[1] in $\mathbb{R}^m$, the sample correlation integral is defined to be the observed proportion of distinct pairs $(\mathbf{x_i}, \mathbf{x_j})$, $1 \leq i < j \leq n$, which are no more than distance ϵ apart. i.e.

$$C_n(\epsilon) = \frac{2}{n(n-1)} \sum_{i \leq j} \mathbf{1}_{\{|\mathbf{x_i} - \mathbf{x_j}| \leq \epsilon\}} \tag{2.7}$$

In the case that $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ are i.i.d. with distribution $\mathcal{F}$, we note that $C_n(\epsilon)$ is the obvious choice for an estimator of the spatial correlation integral $C_{\mathcal{F}}(\epsilon) = \mathcal{F} \times$

[1] We can obtained vectors using delay method because we use a multidimensional space instead of one-dimensional.

$\mathcal{F}(\{(\mathbf{x},\mathbf{y}) \in \mathbb{R}^m \times \mathbb{R}^m |\ |\mathbf{x}-\mathbf{y}| \leq \epsilon\})$. In fact it is trivial to see that $E(C_n(\epsilon)) = C_{\mathcal{F}}(\epsilon)$, and we would to observe *w.p.* 1 convergence of $C_n(\epsilon)$ to $C_{\mathcal{F}}(\epsilon)$ as well as asymptotic normality of suitably rescaled $C_n(\epsilon)$, as $n \to \infty$. This is indeed the case, although the asymptotic do not fall under the trivial i.i.d. format for the sample proportions. The terms in the summation comprising $C_n(\epsilon)$ are not independent, due to the presence of overlapping pairs i.e. pairs $(\mathbf{x_i}, \mathbf{x_j})$ and $(\mathbf{x_i}, \mathbf{x_k})$ that have a member in common. However, we know that $C_n(\epsilon)$ is an U-Statistic and so its convergence behavior can be studied through the theory of U-Statistics[2]. Such an analysis has been carried out by Denker and Keller, who establish the consistency and asymptotic normality of $C_n(\epsilon)$ in the more general case of mixing[3] $\mathbf{x_1}, \mathbf{x_2}, \ldots$ with distribution $\mathcal{F}$.

Given a stationary sequence $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ in $\mathbb{R}^m$, the standard estimation procedure is to plot $\log C_n(\epsilon)$ vs $\log \epsilon$ and carry out a least squares analysis over an ϵ-range where the curve appears linear. The hope is that n is large enough that $C_n(\epsilon) \approx C_{\mathcal{F}}(\epsilon)$ in an ϵ-range where $C_{\mathcal{F}}(\epsilon) \approx \phi\epsilon^{\alpha}$. Obviously this is a double limit problem, since in general we require both $n \to \infty$ and $\epsilon \to 0$.

As an example we can consider a data set of 1000 consecutive iterates from the Henon map. The Hènon map is a discrete-time dynamical system. It is one of the most studied examples of dynamical systems that exhibit chaotic behavior. The Hènon map takes a point (x, y) in the plane and maps it to a new point, for more details related to the map see 69. For the canonical values the Hènon map is chaotic. For other values the map may be chaotic, intermittent, or converge to a periodic orbit. Distances were computed using the square metric.

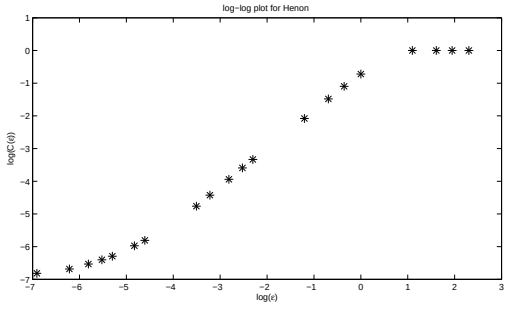


Figure 2.3. Log-log plot for Henon map on $\mathcal{X}$ axis using the square metric.

The example would be considered unusual in practice, because the equation governing the system is known, and we are able to observe the system evolving in its natural

[2]Next sections for more details
[3]For a brief definition of the term see page 16

phase space. The equations associated with real systems are rarely known and data usually arises as a real-valued time series $\{x_i\}$ $i = 1, \ldots, n$, obtained by measuring some component of the system. In such case, the method of time-delay embedding is used. For a positive m, we construct the embedded vectors $\mathbf{x_i} = (x_i, x_{i+1}, \cdots, x_{i+(m-1)})$ $i = 1, \ldots, (n - m + 1)$, in $\mathbb{R}^m$ is then computed, and an estimate $\hat{\alpha}_m$ of the correlation dimension obtained, using the least square method. This procedure is carried out for a sequence of m values, $m = 1, 2, \ldots$ until it is clear that the estimates $\hat{\alpha}_m$ are no longer changing. If the estimates $\hat{\alpha}_m$ do approach a finite asymptotic as m increases, this asymptote is taken as the estimate of the correlation dimension of the underlying system. This procedure, as a whole, is known as the *Grassberger-Procaccia algorithm*.

However, the effect of embedding dimension on estimate quality is relatively minor in the case of the Henon mapping because the Henon attractor is a very low-dimensional object. If the underlying system is high-dimensional (or stochastic), the effect due to increasing embedding dimension rapidly becomes serious - the ϵ-range over which sufficient data is available shrinks quickly. This phenomenon occurs because the components of the embedded vectors act almost independently until m nears the dimension of the underlying system. Using the square metric in $\mathbb{R}^m$, the independence hypothesis implies that the proportion $C_m(\epsilon)$ of m-vectors within ϵ of each other is related to the proportion $C_1(\epsilon)$ of 1-vectors within ϵ of each other by $C_m(\epsilon) = C_1(\epsilon)^m$. This means that, for a given initial sample size, the unstable ϵ-range in $\mathbb{R}^m$ may be very small (or even nonexistent) when m is large. In addition, as it is the smallest ϵ values that are lost, the lower bound of the potential scaling range shifts further and further away from zero as m increases. Since correlation dimension is defined as a limit as $\epsilon \to 0$, this can mean that slopes obtained over the available scaling region are invalid.

An important role is given by estimate method, in this case Grassberger-Procaccia algorithm uses the least squares method. The main issue is to estimate parameters and obtain an accurate variance estimate. It is clear that the variance estimate for the line slope based on least squares methods is not appropriate to show an estimated variance because of the absence of independence distances assumption. Here, we show some results by Cutler related to Grassberger-Procaccia estimator and in particular we focus on the variance computation, considering the dependence of the distances. Let $\mathbf{x_1}, \ldots, \mathbf{x_n}$ be a sequence of random vectors in R^m.The sample correlation integral $C_n(\epsilon)$ takes the form:

$$C_n(\epsilon) = \binom{n}{2}^{-1} \sum_i \sum_{i>j} h(\mathbf{x_i}, \mathbf{x_j})$$

where the function $h(u,v) = 1_{|u-v|\leq\epsilon}$ is symmetric in its arguments (h is called the kernel function[4]). In the case that the original variables $X_1, \ldots, X_N$ are i.i.d the consistency and the asymptotic normality of U-statistics is well known and can be found. These results have been extended by Yoshihara to stationary absolutely regular processes. In Grassberger-Procaccia estimator h function is the indicator function.

Theorem 2. *Suppose $X_1, \ldots, X_N$ is a stationary, absolutely regular sequence in R^m with the marginal distribution $\mathcal{F}$. Let $0 < \epsilon_1 < \epsilon_2 < \ldots < \epsilon_k$ be fixed and let $\boldsymbol{C}_n(\epsilon)$ and $\boldsymbol{C}_{\mathcal{F}}(\epsilon)$ denote the vectors of sample correlation integrals and the real correlation integral for each ϵ-value. Then*

$$\boldsymbol{C}_n(\epsilon) \to \boldsymbol{C}_{\mathcal{F}}(\epsilon) \ w.p.1$$

and

$$\log \boldsymbol{C}_n(\epsilon) \to \log \boldsymbol{C}_{\mathcal{F}}(\epsilon) \ w.p.1 \tag{2.8}$$

there exist nonnegative definite matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ such that

$$\sqrt{n}(\boldsymbol{C}_n(\epsilon) - \boldsymbol{C}_{\mathcal{F}}(\epsilon)) \xrightarrow{d} N_k(0, \boldsymbol{U}) \ as \ n \to \infty$$

and

$$\log \sqrt{n}(\boldsymbol{C}_n(\epsilon) - \log \boldsymbol{C}_{\mathcal{F}}(\epsilon)) \xrightarrow{d} N_k(0, \boldsymbol{V}) \ as \ n \to \infty \tag{2.9}$$

The notation $N_k(0, \mathbf{U})$ denotes a k-variate normal distributions with mean 0 and covariance matrix $\mathbf{U}$, and $\xrightarrow{d}$ indicates convergence in distribution. The components of the matrix $\mathbf{U}$ can also be determined from the U-statistics theory and the matrix

[4]See U-Statistics Theory section

$\mathbf{V} = [v_{ij}]$ is related to the matrix $\mathbf{U} = [u_{ij}]$ by

$$v_{ij} = \frac{u_{ij}}{C_{\mathcal{F}}(\epsilon_i) C_{\mathcal{F}}(\epsilon_j)}$$

If $\mathbf{x_1}, \cdots, \mathbf{x_n}$ are i.i.d. then

$$\begin{aligned} [u_{ij}] &= 4\tau(\epsilon_i, \epsilon_j) = Cov(\mathcal{F}(B(\mathbf{x}, \epsilon_i))\mathcal{F}(B(\mathbf{x}, \epsilon_j))) \\ &= \int \mathcal{F}(B(\mathbf{x}, \epsilon_i))\mathcal{F}(B(\mathbf{x}, \epsilon_j))\mathcal{F}(d\mathbf{x}) - C_{\mathcal{F}}(\epsilon_i) C_{\mathcal{F}}(\epsilon_j) \end{aligned} \tag{2.10}$$

where $\mathcal{F}(B(\mathbf{x}, \epsilon_i))$ is the mass of a random ball of radius ϵ_i. In the special case $\epsilon_i = \epsilon_j$, $\tau(\epsilon_i, \epsilon_j)$ reduces to the variance of the mass in a random ball of radius ϵ_i. If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are correlated but sufficiently mixing $[u_{ij}]$, in particular the sequence has to satisfy the hypothesis of (**Theorem 2**) then variance computation must be modified considering covariance between observations:

$$[u_{ij}] = 4\tau(\epsilon_i, \epsilon_j) + 8 \sum_{m=1}^{\infty} k(m, \epsilon_i, \epsilon_j) \tag{2.11}$$

where $k(m, \epsilon_i, \epsilon_j) = Cov(\mathcal{F}(B(\mathbf{x}_1, \epsilon_i))\mathcal{F}(\mathbf{x}_{1+m}, \epsilon_j))$.

These quantities are estimated by:

$$p(t, \epsilon) = \frac{1}{n-1} \sum_{i \neq t}^{n} 1_{||\mathbf{x}_i - \mathbf{x}_t|| \leq \epsilon} \quad \text{fixed } \mathbf{x_t} \tag{2.12}$$

this represents the proportion of points near $\mathbf{x}_t$ with a distance equal or less than ϵ, so $\hat{\tau}$ and $\hat{k}$ become:

$$\hat{\tau}(\epsilon_i, \epsilon_j) = \frac{1}{n} \sum_{t=1}^{n} p(t, \epsilon_i) p(t, \epsilon_j) - C_n(\epsilon_i) C_n(\epsilon_j)$$

$$\hat{k}(m, \epsilon_i, \epsilon_j) = \frac{1}{n-m} \sum_{i=1}^{n-m} p(t, \epsilon_i) p(t+m, \epsilon_j) - C_n(\epsilon_i) C_n(\epsilon_j)$$

where $C_n(\epsilon_i)$ is the sample correlation integral computed with the indicator function and in this way it is possible to estimate $\hat{\tau}$ and $\hat{k}$ using $p(t, \epsilon_i)$ computed for different ϵ_i.

The correlation dimension estimated for Cutler is the same shown by Grassberger-Procaccia. Least squares method is used. If we consider $x_i = \log \epsilon_i$ and $y_i = \log C_n(\epsilon_i)$ then the slope of ordinary least squares line through data pairs (x_i, y_i), $i = 1, \cdots, k$ is given by

$$\hat{b} = \frac{\sum_{i=1}^{k}(x_i - \overline{x})y_i}{\sum_{i=1}^{k}(x_i - \overline{x})^2} \tag{2.13}$$

Here the model is considered without intercept so $E(y) = 0$ this estimation is possible if the "exact power law" holds so that the log-relation is linear. Now the variance of b can be compute. Since $y_i = y_i(n) \to \log C_{\mathcal{F}}(\epsilon_i)$ as $n \to \infty$, it easy to see that $\hat{b} = \hat{b}(n)$ converges to $\alpha(\epsilon)$ where $\alpha(\epsilon)$ is the slope of the ordinary least squares line through the parameter pairs $x_i = \log \epsilon_i, y_i = \log C_n(\epsilon_i),\ i = 1, \ldots, k$. In general $\alpha(\epsilon) \neq \alpha_{\mathcal{F}}$ due to the lacunar behaviour of $C_{\mathcal{F}}(\epsilon)$.

Using theorem 2 it is possible to obtain the correct standard error of $\hat{b}$ as an estimator of α[5]:

$$\begin{aligned} V(\hat{b}) &= V\left(\frac{\sum_{i=1}^{k}(x_i - \overline{x})y_i}{\sum_{i=1}^{k}(x_i - \overline{x})^2}\right) = \frac{1}{\Big(\sum_{i=1}^{k}(x_i - \overline{x})^2\Big)^2} V\Big(\sum_{i=1}^{k}(x_i - \overline{x})y_i\Big) \\ &= \frac{1}{\Big(\sum_{i=1}^{k}(x_i - \overline{x})^2\Big)} \sum_{i=1}^{k} V\Big((x_i - \overline{x})y_i\Big) \\ &= \frac{1}{\Big(\sum_{i=1}^{k}(x_i - \overline{x})^2\Big)^2} \sum (x_i - \overline{x})(x_j - \overline{x})V(y_i) \\ &= \frac{1}{\Big(\sum_{i=1}^{k}(x_i - \overline{x})^2\Big)^2} \sum (x_i - \overline{x})(x_j - \overline{x})\frac{v_{ij}}{n} \end{aligned} \tag{2.14}$$

An asymptotic 95% confidence interval for α is then given by:

$$[\hat{b} - 2\sqrt{\hat{V}(\hat{b})}; \hat{b} + 2\sqrt{\hat{V}(\hat{b})}] \tag{2.15}$$

[5]In the identification, we delete the m embedding dimension because values should not change for different m, so we can adopt α instead of α_m.

We can remember that the variance estimator by OLS is:

$$V(\hat{b}) = \frac{\sigma^2}{\sum(x_i - \overline{x})^2} \tag{2.16}$$

if σ^2 is unknown an unbiased estimate is given by the sum of errors divided the degrees of freedom:

$$\hat{\sigma}^2 = \frac{\sum(y_i - \overline{y})^2 - \hat{b}^2 \sum(x_i - \overline{x})^2}{n-2} \tag{2.17}$$

It is clear the estimators for the variance are different because (2.16) consider the observation independent but this is false on the other side the estimator (2.14) is based on dependence of observations.

Knowledge of the form of the asymptotic covariance matrix $\mathbf{V}$ make *generalized* least squares an option. Let $\mathbf{X}$ be the $k \times 2$ matrix whose first column consists of all 1's and whose second column is the vector $\mathbf{x}$ where $\mathbf{x}^T = (x_1, \ldots, x_k)$ then the generalized least squares estimate $\hat{b}_*$ of the slope is given by the second entry in the 2 column vectors:

$$\hat{b}_* = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} y \tag{2.18}$$

where $y^T = (y_1, \ldots, y_k)$. It is not clear however, that generalized least squares will be of benefit here. While $\hat{b}_*$ will have reduced variability compared to $\hat{b}$, it will also likely have increased bias in the estimate of the asymptotic quantity $\alpha_{\mathcal{F}}$. This is because the matrix $\mathbf{V}$ will weight in favor of those observations y_i at larger ϵ_i even though it is the smaller ϵ values of $C_{\mathcal{F}}(\epsilon)$ that more accurately produce $\alpha_{\mathcal{F}}$. In a paper Keller and Kunzle [1992] consider the problem in the manner we have just done here. They do however suggest a preliminary chi-square test of the strict linear relationship $\log C_{\mathcal{F}}(\epsilon) = \phi + \alpha_{\mathcal{F}} \log(\epsilon)$ in an ϵ-neighborhood of the origin. If it can be concluded that the parameter pairs $(\log \epsilon, \log C_{\mathcal{F}}(\epsilon))$ are likely to lie on a straight line i.e. that lacunarity is absent or negligible, then the generalized least squares becomes a good choice for estimating $\alpha_{\mathcal{F}}$

2.3 Role of noise

In general the observed signal x_n is obtained from a true trajectory y_n and noise r_n. Observed time series are corrupted by noise and for this reason one of the issue in time series analysis is to reduce noise r_n to know the true signal x_n. Any noise reduction method must assume that the time series to be cleaned can be unambiguously separated into *noise* and *signal* on the basis of some objective criterion.
Conventional methods do not work well with chaotic time series. One has somehow to use the fact that deterministically chaotic (and dissipative) motion takes place on attractors which are smooth sub-manifolds of the total available phase space. This implies that state vectors constructed from delay variables are constrained to fall onto geometrical objects which are locally linear.

When adding noise to an otherwise deterministic system, we have to distinguish between *dynamic* and *measurement* noise. Assume that the noise-free dynamics would be:

$$y_{n+1} = f(y_n) \tag{2.19}$$

We define measurement noise if there exists a trajectory satisfying this exact dynamics, but the measured trajectory is corrupted by additive noise,

$$x_n = y_n + r_n \tag{2.20}$$

Dynamic noise, in contrast, is added already during the evolution,

$$x_{n+1} = f(x_n) + r_n \tag{2.21}$$

so that no near-by trajectory satisfying the exact dynamics needs to exist a priori. The *shadowing problem* deals just with the question whether such a trajectory does exist and how to find it. It is clear the shadowing problem is harder than the problem of

removing measurement noise.

The characterization of nonlinear time series in the presence of measurement noise is a problem of great current interest. For the ideal case of noise-free deterministic time series, the reconstruction theorem has led to a number of powerful characterization methods (here we will use the method of delay time). Measurement noise, however, is known to put severe limitations on the estimation of dynamical invariants from time series with these methods.

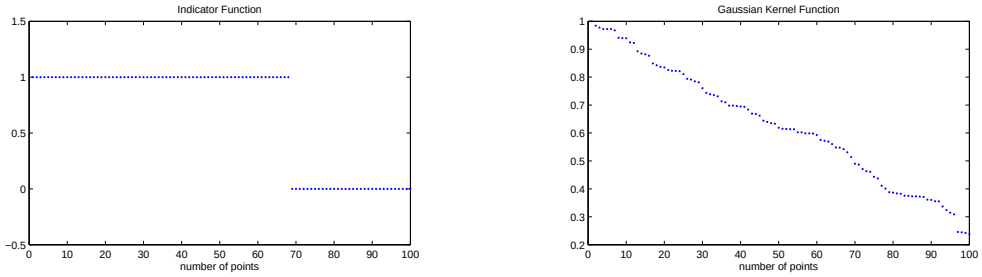
The most famous method used to characterized noise-free time series is the Grassberger-Procaccia estimator studied in the last paragraph.

Basically two approaches can be distinguished concerning the analysis of time series with measurement noise. The first is to separate the noise and the underlying time series with a noise reduction method see [Kostelich and Schriber, 1993] and [Grassberger et al., 1993]. The second is based on characterizing the modified delay vector distribution. By calculating the effect of noise on the correlation integral [Shouten et al., 1994], obtained a method for estimating the correlation dimension in the case of bounded independent, identically distributed (IID) noise. Schriber [1993] has proposed a method for estimating the noise level of a deterministic time series contaminated with unbounded IID Gaussian measurement noise. The effect of this noise on the correlation integral has also been investigated by see [Smith, 1992], who used an approximation of the correlation integral to estimate the correlation dimension for small noise levels. The analytic difficulties which prevent the estimation of invariants at higher noise levels appear to be related to the contrast between the smooth Gaussian noise distribution on the one hand and the abrupt nature of the kernel function in the correlation integral on the other.

2.4 Gaussian Correlation Integral

In the 1996 Diks proposed an estimator for the correlation integrals. The basic idea of this new estimator was the same to the $C(\epsilon)$ but in that view there is a difference about the weights to give at all inter-point distances pairs. In particular the weight system follows the gaussian distribution. In this way pairs (x_i, x_j) which have a small distances weight more than those with large distances. So the contrast between the Gaussian noise and the kernel function should be go over. In Figure 2.4(a) it is possible to see the main differences between two weight systems. Using the indicator function a precise and defined cut-off is visible so distances less than ϵ are considered. The Gaussian system considers all pairs distances given several weight in this sense it is

impossible to find a precise cut-off, as figure shows.



(a) (b)

Figure 2.4. Different systems weight. (a) Indicator function and (b) normalized Gaussian function

The correlation integral defined by Grassberger-Procaccia can be generalized to

$$T_m(\epsilon) = \int\int w(|\mathbf{x}-\mathbf{y}|/\epsilon) f_m(\mathbf{x}) f_m(\mathbf{y}) d\mathbf{x} d\mathbf{y} \tag{2.22}$$

where $w(\bullet)$ is a kernel function. The parameter ϵ will be referred to as the bandwidth. Using the Gaussian kernel function

$$w(x) = e^{-x^2/4} \tag{2.23}$$

a version of the correlation integral,

$$T_m(\epsilon) = \int\int exp\left\{-\frac{|\mathbf{x}-\mathbf{y}|^2}{4\epsilon^2}\right\} f_m(\mathbf{x}) f_m(\mathbf{y}) d\mathbf{x} d\mathbf{y} \tag{2.24}$$

is obtained which will be referred to as the Gaussian kernel correlation integral; in particular when we use $T_m(\epsilon)$ we consider the Gaussian correlation integral otherwise in case of Grassberger-Procaccia correlation integral we adopt $C(\epsilon)$.

Now, we are interested to show the model of $T_m(\epsilon)$ in the noise-free case. At first, we take m fixed and consider a deterministic time series with correlation dimension α, then the scaling law

$$T(\epsilon) \sim \epsilon^{\alpha} \;\; \epsilon \to 0 \;\; \text{for m fixed} \tag{2.25}$$

More generally, any kernel $w(x)$ which decreases monotonically in x for $x \geq 0$ and for

which $\lim_{\epsilon\to 0}\epsilon^{-p}w(x/\epsilon)=0$ pointwise for $x>0$ and for any $p\geq 0$ implies the scaling law (2.25). It is possible to show the m-dependence of (2.25) by expressing $T_m(\epsilon)$ as:

$$T_m(\epsilon)=\int w(r/\epsilon)\rho_m(r)dr \tag{2.26}$$

where $\rho_m(r)=\frac{dC_m(r)}{dr}$ is the distribution of the inter-point distances r. It is shown that the correlation integral calculated with the Euclidean norm behaves as $C_m(\epsilon)\sim e^{-K\tau m}(r/\sqrt{m})^{\alpha}$ for $r\to 0,\ m\to\infty$ which implies $\rho_m(r)\sim e^{-K\tau m}m^{-\alpha/2}$ for fixed r. The factor $e^{-K\tau m}m^{-\alpha/2}$ describes the m dependence. We therefore find

$$T_m(\epsilon)\sim e^{-K\tau m}(r/\sqrt{m})^{\alpha}\epsilon^{\alpha}\text{ for }r\to 0,\ \ m\to\infty \tag{2.27}$$

for Gaussian correlation integral in the noise-free case with the Euclidean norm. We could remove the factor $m^{-\alpha/2}$ in (2.27) by defining and m-dependent bandwidth. There is, however, a practical reason for not using this freedom and proceeding with (2.27). Due to the finiteness of the attractor there usually is an upper bandwidth up to which the behavior (2.27) is observed, and it is approximately independent of m.

A Gaussian kernel member is picked from this class and it is possible to derive analytically its behaviour in presence of Gaussian measurement noise of this form $N(0,\sigma^2)$. The new estimator becomes:

$$T_m(\epsilon)=\int\int\exp\left\{-\frac{|\mathbf{x}-\mathbf{y}|^2}{(4\epsilon^2+4\sigma^2)}\right\}f_m(\mathbf{x})f_m(\mathbf{y})d\mathbf{x}d\mathbf{y} \tag{2.28}$$

here σ represents the measurement (additive) noise and f is the underlying noise-free distribution. The behaviour of the double integral is found from the definition given in (2.24) together with the noise-free scaling law (2.5), leading to

$$T_m(\epsilon)\cong\phi\left(\frac{\epsilon^2}{\epsilon^2+\sigma^2}\right)^{m/2}e^{-K\tau m}m^{-\alpha/2}\sqrt{\epsilon^2+\sigma^2}^{\alpha} \tag{2.29}$$

where K is the correlation entropy define as $K=\tau^{-1}\lim_{\epsilon\to 0}\lim_{m\to\infty}\frac{-\log C_m(\epsilon)}{m}$ and ϕ is a normalization constant.

In practice, the standard deviation σ of the noise level is fixed at a nonzero value. We are not able to let $\sqrt{\epsilon^2+\sigma^2}$ go to zero. Nevertheless, we expect relation (2.29) to hold good in a range of small values in ϵ if the noise level σ is not too large. For small

values of ϵ and m fixed the Gaussian correlation integral behaves as $T_m(\epsilon) \sim \epsilon^m$ which is a manifestation of the m-dimensionality of the set of noisy delay vectors. Taking the limit $\sigma \to 0$ on the other hand, gives back the scaling relation of noise-free case:

$$T_m(\epsilon) \sim e^{-Km} m^{-\alpha/2} \epsilon^{\alpha} \quad \text{for } \epsilon \to 0,\ m \to \infty \tag{2.30}$$

The Gaussian kernel correlation integrals $T_m(\epsilon)$ can be consistently estimated by replacing the integrals over the delay vectors distributions in (2.28) with an average over delay vectors are assumed to be independently distributed according to $f_m(\epsilon)$. The estimate $T_m(\epsilon)$ becomes

$$\hat{T}_m(\epsilon) = \frac{1}{n(n-1)} \sum_i \sum_{i \neq j} \exp\left[-\frac{|\mathbf{x_i} - \mathbf{x_j}|^2}{4\epsilon^2} \right] \tag{2.31}$$

$$= \frac{1}{n(n-1)} \sum_i \sum_{i \neq j} \psi_{i,j}(\epsilon) \tag{2.32}$$

For different values of the noise level, a Marquardt non-linear fit procedure for the parameters ϕ, α, D and σ is used. For each m, the values $\hat{T}_m(\epsilon)$ and $\hat{T}_{m+1}(\epsilon)$ were fitted simultaneously to the model function (2.29). The standard deviations of the estimates $\hat{T}_m(\epsilon)$ are taken as the weights in the fit procedure. Assuming independence of the distances, the variances $\mathcal{V}(\hat{T}_m(\epsilon))$ is estimated as:

$$\mathcal{V}(\hat{T}_m(\epsilon)) = \frac{1}{n(n-1)} [\overline{\psi^2_{i,j}}(\epsilon) - \overline{\psi_{i,j}}(\epsilon)^2] \tag{2.33}$$

where the bars denote average over the pairs (i, j).
The main assumption is the independence of distances but this is not reasonable assumption because the pair distances are dependent because (i, j) and (i, k) are correlated, so it will be needed to use a different variance formula.

2.5 Takens estimator

We now examine a maximum likelihood approach to obtain dimension estimates. This approach is based on the exact power law model 2.5. This model features two unknown

parameters ϕ and α and we are interested in α.

Let $D_1, D_2, \ldots, D_n$ be i.i.d. nonnegative real-valued random variables with distribution function $\mathcal{F}$ and suppose that $\mathcal{F}$ follows an exact power law for small ϵ,

$$\mathcal{F}(\epsilon) = \phi\epsilon^{\alpha} \quad \text{for } 0 < \epsilon \leq \epsilon_0 \tag{2.34}$$

The conditional distribution of D_j, given that $D_j \leq \epsilon_0$ is then

$$\mathcal{F}(d|\epsilon) = \left(\frac{d}{\epsilon_0}\right)^{\alpha} \quad \text{for } 0 < d \leq \epsilon_0 \tag{2.35}$$

which has probability density function

$$f(\epsilon) = \alpha\epsilon_0^{-\alpha} d^{\alpha-1} \quad \text{for } 0 < d \leq \epsilon_0 \tag{2.36}$$

Note that the parameter ϕ has disappeared from the conditional model, leaving only α, the dimension quantity of interest. Therefore, if we select from the original sample only those D_j's satisfying $D_j \leq \epsilon_0$ (for convenience we will label these as $D_1, \ldots, D_t$ and assume that $t \geq 1$) then the likelihood of this selected sub-sample is given by

$$L(d_1, \ldots, d_t; \alpha) = \alpha^t \epsilon_0^{-t\alpha} (d_1, \ldots, d_t)^{\alpha-1} \tag{2.37}$$

Maximizing $L(d_1, \ldots, d_t; \alpha)$ with respect to α is equivalent to solving the likelihood equation

$$\frac{t}{\alpha} - t\alpha \log \epsilon_0 + (\alpha - 1) \sum_{i=1}^{t} \log d_i \tag{2.38}$$

which yields the maximum likelihood estimate (MLE):

$$\hat{\alpha} = \frac{t}{\sum_{i=1}^{t} \log(\epsilon_0/d_i)} \tag{2.39}$$

This estimator is often called Takens estimator because he was first noted that the method of maximum likelihood could be applied to the estimation of correlation dimension. If some suitable transformation can be applied to the original observations $X_1, \ldots, X_n$ to produce random variables $D_1, \ldots, D_t$ which are i.i.d. with distribution

$H(\epsilon)$, then we may consider using maximum likelihood, with $H(\epsilon)$ in the role of $F(\epsilon)$. Of course an exact power law generally does not hold for fractal data, and the effect of lacunarity on the quality and consistency of maximum likelihood estimates is a point of considerable interest. In 1985 Takens initially considered maximum likelihood in the context of correlation dimension under two working assumptions. The first working assumption is that the scaling relation is $C_{\mathcal{F}}(\epsilon) = \mathcal{F} \times \mathcal{F}(|\mathbf{x} - \mathbf{y}| \leq \epsilon) = \phi(\epsilon)\epsilon^{\alpha}$ again a distribution function in ϵ. In this case, given i.i.d. observations $X_1, \ldots, X_n$ from $\mathcal{F}$, we compute $t = \binom{n}{2}$ pairwise distances $D_{ij} = |X_i - X_j|$. The second working assumptions is to consider the independence of observations: of course the D_{ij}'s are not i.i.d., due to the statistical dependence between pairs (X_i, X_j) and (X_i, X_k) having a component X_i in common. However for a small radius of ϵ_0 , the problem of overlapping pairs is relatively insignificant, so in many cases it may be reasonable to treat the D_{ij} as i.i.d. Lacunarity is also a factor in the behaviour of the correlation integral, although the fact that $C_{\mathcal{F}}(\epsilon)$ is a global quantity averaged over all points in the phase space mitigates this effect to some degree.

In our case, the α estimator for the correlation dimension is

$$\hat{\alpha}_T = -\left(\frac{2}{n(n-1)} \sum_{i<j} \log \frac{|\mathbf{x_i} - \mathbf{x_j}|}{\epsilon_0} \right)^{-1} \tag{2.40}$$

and it is called Takens estimator for correlation dimension. The central part is composed by a kernel function made in that form:

$$h(x_i, x_j) = \sum_{i \neq j} \log \frac{|x_i - x_j|}{\epsilon_0} \tag{2.41}$$

The variance of the estimate in the limit normal law can be expressed through the Fisher information:

$$I_F(\alpha_T) = E\left(\frac{\partial \ln L}{\partial \alpha} \right)^2 \tag{2.42}$$

where L is the likelihood function defined as (2.37). The quantity represents the information contents about the parameter of interested α when we look at the sample of the

distances $d_1, d_2, \ldots, d_k$. The asymptotic variance of $\hat{\alpha}$ is then:

$$\mathrm{Var}_\infty(\hat{\alpha}_T) = \frac{1}{I_F(\alpha_T)} = \frac{\alpha_T^2}{k} \tag{2.43}$$

Note that the value of ϵ_0 should be taken small enough in order to validity of (2.40). The expression of the (2.43) will change in presence of dependent distances $\{d_i\}$.

***

It is possible to look at the Takens estimator as a tail estimator, in fact if a power law hold and if the sequence of random variables have the same marginal distribution function $\mathcal{F}$ and if the $\overline{\mathcal{F}} = 1 - \mathcal{F}$ is regularly varying at ∞, ,namely exists an $\alpha > 0$, such that

$$\begin{aligned} \overline{\mathcal{F}}(tx)/\overline{\mathcal{F}}(x) &\to t^{-\alpha} \qquad \text{as } x \to \infty \text{ for all } t > 0 \\ \overline{\mathcal{F}}(x) &= x^{-\alpha}\phi(x) \quad x > 0 \text{ for some slowly varying function } \phi. \end{aligned} \tag{2.44}$$

The class of the distributions having this behavior, called tail behavior, is infinitely large, and it is known to coincide with the maximum domain of attraction of the extreme value distribution $\exp(-x^{-\alpha}),\ x > 0$.

What we want to investigate is the estimate of $-\alpha$, called regular variation index, when observing $X_1, \ldots, X_n$. It is intuitively clear that if little or no additional structural information on $\mathcal{F}$ is available, which we assume to be the case, any inference on α should be made with the tail portion of the empirical distribution of the sample. For $1 \leq j \leq n$, write $X_{(j)} = X_{(n:j)}$ for the j-th largest value of $X_1, \ldots, X_n$, for $x > 0$ let be $x^* = \log x$. Hill Hill [1975] was first to propose an estimator for $-\alpha$:

$$H_n = m^{-1} \sum_{j=1}^{m} X^*_{(j)} - X^*_{(m+1)} \tag{2.45}$$

Asymptotic proprieties of H_n, including consistency and asymptotic normality, were studied by letting m vary with n such that

$$m \to \infty \text{ and } m/n \to 0 \text{ as } n \to \infty.$$

This estimator does not work well in presence of depending data, but it is possible to

introduce a little change to obtain an estimator which converges to the Hill's estimator [Hsing, 1991].

The tail index is well linked with the correlation dimension because of the presence of region where the scaling law (2.5) holds. Here, it is possible to define the distances set such that $|x_i - x_j| = \delta_{(ij)}$ and ordering $\delta_{(n:1)}, \delta_{(n:2)}, \ldots, \delta_{(n:n)}$ and let be $\delta^*_{(n:j)} = \log \delta_{(n:j)}$. The Hill's estimator (2.45) is:

$$H_n = \frac{1}{m} \sum_{j=1}^{m} \delta^*_{(n:j)} - \delta^*_{(n:m+1)} \tag{2.46}$$

Because $\delta^*_{(n:m+1)}$ is fixed so we can consider it as $\log \epsilon$ so H_n:

$$\begin{aligned} H_n = & \quad \tfrac{1}{m} \textstyle\sum_{j=1}^{m} \log \delta_{(n:j)} - \log \epsilon \\ = & \quad \tfrac{2}{n(n-1)} \textstyle\sum_{i,j} \log \tfrac{|x_i - x_j|}{\epsilon} \end{aligned} \tag{2.47}$$

The estimator (2.47) looks like to the Takens estimator for the correlation dimension (2.40).

2.6 The Beta-Binomial Estimator

If the correlation integral satisfies an exact power law in some neighborhood $(0, \epsilon_0)$ of the origin, then assuming the independence of distances, Takens maximum likelihood method is the optimal estimation procedure and leads to accurate asymptotic confidence intervals. However, if an exact power law fails to hold, the MLE can perform poorly and may produce invalid confidence intervals. The presence of lacunarity in the correlation integral may make these intervals inapplicable to $\alpha_{\mathcal{F}}$. It would be desirable to have an estimation method that could detect, and compensate for, the presence of lacunarity in the underlying model, yet perform nearly as well as Takens' procedure when in fact an exact power law does not.

Smith has suggested the use of *beta-binomial mixture model* to achieve this goal. Such mixtures are used frequently in statistics to compensate for possible extra variation in a model. We begin by setting the notation and showing how the problem can be cast in the form of binomial model.

We will write

$$C_{\mathcal{F}}(\epsilon) = \phi(\epsilon)\epsilon^{\alpha_{\mathcal{F}}} \tag{2.48}$$

where, in general, $\phi(\epsilon)$ is an oscillating function. We select a sequence $0 < \epsilon_k < \ldots < \epsilon_1 < \epsilon_0$ of ϵ values according to the geometric ration $\epsilon_j = \epsilon_0\gamma^j$ for $j = 1, \ldots, k$, where $0 < \gamma < 1$. At each ϵ_j we count the observed number N_j of distinct pairs (X_i, X_j) satisfying $|X_i - X_l| \leq \epsilon_j$. We then note that N_{j+1} given N_j has a binomial distribution with parameters N_j and $p_{j+1} = \frac{C_{\mathcal{F}}(\epsilon_{j+1})}{C_{\mathcal{F}}(\epsilon_j)}$ i.e.

$$P(N_{j+1} = n|N_j) = \binom{N_j}{n} p_{j+1}^n (1 - p_{j+1})^{N_j - n}$$

for $n = 0, 1, \ldots, N_j$. Applying the scaling law (2.48) and the identity $\epsilon_j = \epsilon_0\gamma^j$, we see that in fact:

$$p_j = \frac{\phi(\epsilon_j)}{\phi(eps_{j-1})}\gamma^{\alpha_{\mathcal{F}}} \tag{2.49}$$

If an exact power law holds aver $(0, \epsilon_0)$ that means that $\phi(\epsilon)$ may be replaced by a constant ϕ in (2.48) then we obtain a binomial model with $p_j = p = \gamma^{\alpha_{\mathcal{F}}}$ for all j. In this case, the common value p can be estimated by maximum likelihood. The likelihood function

$$L(N_1, N_2, \ldots, N_k|N_0; p) = P(N_k|N_{k-1})P(N_{k-1}|N_{k-2}) \ldots P(N_1|N_0) \tag{2.50}$$

that is proportional to

$$p^{(N_k + N_{k-1} + \ldots N_1)}(1 - p)^{(N_0 - N_k)} = p^{\sum_{j=1}^k N_j}(1 - p)^{(N_0 - N_k)} \tag{2.51}$$

The log-likelihood becomes

$$\log L(N_1, N_2, \ldots, N_k|N_0; p) \propto \sum_{j=1}^{k} N_j \log p + (N_0 - N_k)\log(1 - p) \tag{2.52}$$

Maximizing (2.52) with respect to p, the MLE estimate $\hat{p} = \frac{\sum_{j=1}^k N_j}{\sum_{j=0}^{k-1} N_j}$. From the identity

$p = \gamma^{\alpha_{\mathcal{F}}}$ we then obtain the estimate $\hat{\alpha} = \frac{\log \hat{p}}{\log \gamma}$. While this estimate is not quite as optimal as the Takens MLE, Smith points out that $\hat{\alpha}$ performs nearly as well as Takens estimator when k (the number of ϵ values) is large and γ is close to 1. This can be understood by observing that Takens estimator uses detail information from every single observation; that above discrete binomial method does not, but when k is large enough and γ is close to 1, the resulting ϵ-mesh is so fine that we essentially recover the individual observation.

In the general lacunar case, however, the p_j's are distinct quantities, and the differences among them will introduce added variability into the data. In the beta-binomial model, the p_j's are treated as i.i.d. beta random variables with mean p. The probability density function of a beta variable has two parameters ($\beta > 0$ and $\delta > 0$) and is given by

$$f(p_j; \beta, \delta) = \frac{\Gamma(\beta + \delta)}{\Gamma(\beta)\Gamma(\delta)} p_j^{\beta-1}(1 - p_j)^{\delta-1} \text{ for } 0 < p_j < 1 \tag{2.53}$$

Now the conditional likelihood of $N_1, \ldots, N_k$ given N_0 and $p_1, \ldots, p_k$ is proportional to

$$\prod_{j=1}^{k} p_j^{N_j}(1 - p_j)^{(N_{j-1} - N_j)} \tag{2.54}$$

The unconditional likelihood is then obtained by integrating out $p_1, \ldots, p_k$ from (2.54), leaving only the dependence on the underlying parameters β and δ i.e.

$$\begin{aligned} L(N_1, N_2, \ldots, N_k | N_0; \beta, \delta) = &\int_0^1 \ldots \int_0^1 \prod_{j=1}^{k} p_j^{N_j}(1 - p_j)^{(N_{j-1} - N_j)} \\ &\times \prod_{j=1}^{k} f(p_j; \beta, \delta) dp_1 \ldots dp_k \end{aligned} \tag{2.55}$$

A prime reason for choosing the beta distribution is that the integration in (2.55) can be accomplished in closed form. The resulting likelihood is a function of β and δ and can be maximized numerically to obtain the estimated $\hat{\beta}$ and $\hat{\delta}$. The mean of the beta distribution is given by $\beta/(\beta + \delta)$, so the estimate of the average value $p = E(p_j)$ becomes $\hat{\beta}/(\hat{\beta} + \hat{\delta})$. We then again take $\hat{\alpha} = \frac{\log \hat{p}}{\log \gamma}$. From the beta model, the variance of p_j is given by $p(1-p)\tau$ where $\tau = (\beta + \delta + 1)^{-1}$ represents the additional variability in the model due to the variations among the p_j's. On the basis of preliminary results, the

beta-binomial estimator appears to perform well in most situation. In the lacunar case its bias remains fairly small, and wider confidence intervals are produced to compensate for the added variability. When an exact power law does hold, τ is generally estimated close to 0, and the results approximate those of the binomial model. However the beta-binomial estimator does not work well in the presence of noise.

For other estimator see Theiler and Lookman [1993] for the chord estimator and the Ellner's estimator.

Chapter 3

An open question: choice of the scaling region

One of the most important open question is the choice of scaling region over which the estimate of the correlation dimension is done. This represents an unsolved and open question, because *"of course the procedure is not entirely objective, and estimation of the correlation dimension α is often as much an art as it is a science"* pag. 21 in Diks [1999]

Only if the scaling law holds then it is possible to find an unbiased estimate for the correlation dimension.

When the correlation dimension is estimated using the (2.4) on a finite set, the choice of ϵ plays an important role. Typical log-log plots of $\hat{C}(\epsilon)$ against ϵ exhibit the following pattern: at large values of ϵ the slope is nearly horizontal; at medium values of ϵ there is a range of ϵ for which the slope is constant; and for small values of ϵ the graph gets very ragged with a sharp break downwards. The slope of the graph in the middle range of ϵ's is used as an estimate of α. When we use the formula (2.4) we introduce a bias in the estimate of the dimension of the data set.

There are a lot of studies and proposals to find the "true" scaling region over that it is possible to estimate the correlation dimension α. In particular here we will consider just two or three of them. The first method takes in consideration proprieties of U-Statistics, the second starts from the BDS test used in the independence hypothesis testing and the third is based on the chi-square test of the scaling region search. Aim of the studies is the same: find the range of the values for ϵ because a lot of empirical

analysis were made with important outcomes. In particular in Kocenda and Briatka [2005] we can find a complete reference about the ϵ values choice.

3.1 Some considerations

It is important to remember some guidelines as we can find in Diks [2004]. It is known from chaotic time series analysis that estimates of the correlation dimension can vary with the embedding dimension and the selected range of ϵ values. Examining the estimated correlation dimension as a function of m is often informative. Estimates which do not saturate with m indicate that either the underlying assumptions are wrong, the scaling region used is not appropriate or, the number of data are insufficient.

In the choice of the scaling region there are two practical limitations. First of all, the theoretical or model correlation integral is derived for small ϵ values, which indicates that one might expect an upper length scale above which this model performs poorly. In fact, for sufficiently large ϵ the model predicts correlation integrals which are larger than one, which is impossible. On the other hand, statistical fluctuations dominate the estimated when the correlation integral is small. This puts a practical lower limit on the values of the estimated correlation integrals that provide useful information. For the traditional hard kernel correlation integral do not pose a problem. There a natural lower limit is set on correlation integral, equal to one over the number of distances among all points, the smallest possible nonzero value of the correlation integral. For the Gaussian kernel such a natural lower bound does not exist. However, it is clear that small estimates of the correlation integral might suffer from large relative statistical errors.

Dechert [2003] proposed a systematic way of choosing ϵ which is based on the statistical proprieties of U-Statistics. It is a method for choosing a sequence of $\{\epsilon_n\}$ such that $\epsilon_n \to 0$ and for which some particular statistical results still hold.

The consequence of theorem 8 in [Dechert, 2003] is that if ϵ_n converges slowly enough to 0 so that a quantity[1] diverges , then $\lim_{n\to\infty} |F_n(\cdot;\epsilon) - \Phi|_\infty = 0$ i.e. the empirical distribution function F_n converges to the Normal distribution function Φ.

[1] Let be $C_n(\epsilon_1,\epsilon_2) = \frac{2}{n(n-1)} \sum\sum_{1\le<i\neq<j\le n} \mathbf{1}_{\{|x_i-x_j|\le\epsilon_1\}}\mathbf{1}_{\{|x_i-x_j|\le\epsilon_2\}}$, then a new quantity is define

$$S_n = C_n(\epsilon_1,\epsilon_2) - C_n(\epsilon_1)C_n(\epsilon_2)$$

and for $n \to 0$ converges to 0.

Let be $K(\epsilon) = E[(F(X_1+\epsilon) - F(X_1-\epsilon))^2]$, if $K(\epsilon_i) - C(\epsilon_i)^2 > 0$ for $o = 1, 2$ and if random variables $\{X_t\}$ are independent then,

In particular if data $\{x_n\}$ are corrupted by noise $\{r_n\}$ as follows:

$$\begin{aligned} y_{n+1} &= f(y_n) \\ x_n &= y_n + \zeta r_n \end{aligned} \tag{3.1}$$

where $\{r_n\} \sim IID(0,1)$ then for a large ζ one would expect that the sequence $\{x_n\}$ would behave like a random sequence and that $S_n(\epsilon, k^{1+c+\ldots+c^{m-1}}\epsilon^{c^m})$ would converge to zero. For small ζ one would expect that the sequence would behave like deterministic data and that $S_n(\epsilon, k^{1+c+\ldots+c^{m-1}}\epsilon^{c^m})$ would converge to a positive constant, where c and k are two constant.

The second approach is based on the use of BDS test which uses a range for hypothesis testing. This test is a nonparametric method of testing for nonlinear patterns in time series. It is based on the Correlation Integral and it is unique in ability to detect nonlinearities independent of linear dependencies in the data. The null hypothesis is that data in a time series are independently and identically distributed (i.i.d.) and an alternative is not specified. In order to conduct the BDS test, two free variables - m embedding dimension and ϵ values of the scaling region - must be chosen *ex ante*. We know that chaotic system of low dimensionality can generate seemingly random numbers that may give an impression of white noise, thereby hiding their true nature. Under presumed randomness, a nonlinear pattern can hide without detected. Detection of nonlinear hidden pattern in such time series provides important information about their behaviour and improves forecasting ability over short time periods.

Under the null hypothesis data are i.i.d. the correlation dimension given by the slope estimate on log-log line, computed at each m value, should be equal to the respective embedding dimension m i.e. $\alpha_m = m$.

In the choice of the range we have to consider this is closely related to the composition of the analyze data. Therefore it is possible that one range is more appropriate for one kind of data and a different range for another one. In view of this situation some authors found that the power and the size of the BDS test is maximized when ϵ is chosen between 0.5 and 1.5 of the standard deviation of the sample using proportional increments. In Kocenda and Briatka [2005] they propose to use the interval $(0.6\sigma - 1.90\sigma)$ as optimal one, because over other intervals we can obtain biased estimator, where σ

$$\sqrt{n}\frac{S_n(\epsilon_1,\epsilon_2)}{\sigma(\epsilon_1,\epsilon_2)} \xrightarrow{D} N(0,1)$$

where $\sigma(\epsilon_1,\epsilon_2)$ is a function on $K(\epsilon)$ and $C(\epsilon)$. The quantity found is $\frac{n[K(\epsilon_n)-C(\epsilon_n)^2]^3}{C(\epsilon_n)^2}$.

represents the mean square error of the system.

The last method is based on the maximum likelihood principle used by Takens; t is possible to derive the explicit expressions for the maximum likelihood estimate of α and its asymptotic variance.

In the real setting we have to give up the pure power law, so a procedure to check the presence of the scaling law is needed. One solution to do it is the use of χ^2-test based on a quantitative measure of validity of the power law distribution hypothesis.

Suppose for a given ϵ_0 we have R non-overlapping independent pairs of points, the distance within each pair being less than ϵ_0. We can divide the interval of presumed scaling $[0, \epsilon_0]$ into k subintervals $[a_0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k)$, where $a_0 = 0, a_k = \epsilon_0$. The probability p_i that a distance between two randomly chosen points lies within the given interval is in according to (2.35) such that

$$p_1 = \left(\frac{a_1}{\epsilon_0}\right)^{\alpha}, \quad p_2 = \left(\frac{a_2}{\epsilon_0}\right)^{\alpha} - \left(\frac{a_1}{\epsilon_0}\right)^{\alpha}, \quad \ldots \quad p_t = 1 - \left(\frac{a_{t-1}}{\epsilon_0}\right)^{\alpha} \tag{3.2}$$

In a general interval $[a_{j-1}, a_j)$ let m_j be the number of the distances lying in, so that $\sum_{j=1}^{k} = R$. Then the χ^2-statistics is

$$\chi^2 = \sum_{j=1}^{k} \frac{(m_j - Rp_j)^2}{Rp_j} \tag{3.3}$$

In the limit $N \to \infty$ this statistics has the χ^2-distribution with $k-2$ degrees of freedom, since the number of fitting parameters is one. The number of the sub-intervals should be not too large, especially for small R. Using the test theory we compare the empirical value (3.3) with a theoretic one depending on the significance level λ. We accept the hypothesis of behaviour such that in (2.35) when $\chi^2 \leq \chi^2_\lambda$ and reject it otherwise.

We can note the estimate of α must precede this test. On the other hand the χ^2-statistics and the probability associated with it can be used as the indicator in the choice of the upper cutoff of the scaling range ϵ_0.

The choice of ϵ_0 can be made as follows. We can start with some initial value ϵ_1 which is small enough, but at the same time the total number of distances R within the radius ϵ_1 must be sufficient to calculate the statistics. For this value of the upper cutoff of the scaling range ϵ_1 we compute $\hat{\alpha}(\epsilon_1)$ and $\chi^2(\epsilon_1)$. We can increase the value of the cutoff and we can verify each time the power law distribution hypothesis with the help of χ^2-test. Finally, we choose as ϵ_0 the maximal cutoff ϵ_i for which the hypothesis of

the distribution (2.35) is still accepted.

There are some practical problems related to the χ^2-test the most important are the estimate's determination of α which the unknown quantity to estimate, the absence of a criterion to choose the initial value of ϵ. It is important to remember some features of ϵ in particular the set of ϵ_i must be large enough to have the possibility to compute distances, at the end the chosen value ϵ_0 must be the smallest value such that power law hold.

Chapter 4

U-Statistics Theory

The U-Statistics theory have been introduced independently by Halmos (1946), Hoeffding (1948) and von-Mises introduced von-Mises-Statistics (1947). The reason to introduce these new Statistics in each of these papers was rather different.

In general U-Statistics theory is quite-well known, so here we will describe in short some basic elements. First of all we will start from definition, then examples are proposed and the last step is about basic theorems. This chapter represents a tool chapter because we will start from these considerations and results to expand our work in the following sections.

4.1 Definition and Examples

It is possible to have a definition for U-Statistics considering $(X_n)_{n\in N}$ as a stationary sequence of random vectors with marginal distribution function F, taking in R^m or a subset of it. Hoeffding initially introduced U-Statistics as estimator for functionals of the form:

$$\theta(F) = \int_{R^m} h(x_1, \ldots, x_m) dF(x_1) \ldots dF(x_m), \tag{4.1}$$

where the kernel function $h : \mathbb{R}^m \longrightarrow \mathbb{R}$ is symmetric in its arguments. Hoeffding showed that, in case $X_1, X_2 \ldots$ are independent and the distribution F is completely

unknown, the minimum variance unbiased estimator of $\theta(F)$ is given by

$$U_n(h) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \ldots < i_m \le n} h(X_{i_1} \ldots X_{i_m}). \tag{4.2}$$

$U_n(h)$ is called the *U-Statistics of degree m* corresponding to the kernel h.

All proprieties hold a generic m, in this work we show the case for $m = 2$ because this is the degree used. Here we compare two similar statistics *U and von-Mises Statistics.*

Definition 2 (*U and von-Mises Statistics* when $m = 2$). *Let $h : R^2 \longrightarrow R$ be a measurable symmetric function, i.e. $h(x, y) = h(y, x)$ for all $x, y \in R^2$. We then define the U-Statistic U_n in case of $m = 2$ by*

$$U_n(h) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} h(X_i, X_j) \tag{4.3}$$

and the von-Mises-Statistic $V_n(h)$ by

$$V_n(h) = \frac{1}{n^2} \sum_{1 \le i < j \le n} h(X_i, X_j). \tag{4.4}$$

h is the kernel function of both statistics.

We can easily extend the definition of U-Statistics and von-Mises-Statistics to kernels such that $h : \mathbb{R}^m \longrightarrow \mathbb{R}$ in which case $U_n(h)$ and $V_n(h)$ are defined as averages of $h(X_{i1}, \cdots, X_{im})$.

By symmetry of kernel function h we can rewrite the U-Statistics as

$$U_n(h) = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} h(X_i, X_j) \tag{4.5}$$

The essential difference between U-Statistics and von-Mises thus lies in the fact the diagonal terms $h(X_i, X_j)$ are included in the von Mises-Statistics and excluded in the

U-Statistics. In fact the sum of diagonal terms is defined as follows

$$n^2 V_n(h) - n(n-1)U_n(h) = \sum_{i=1}^{n} h(X_i, X_i) \tag{4.6}$$

and since the asymptotic behavior of the partial sum $\sum_{i=1}^{n} h(X_i, X_i)$ is well understood, one can fairly easily obtain results for U-Statistics from corresponding results for von-Mises-Statistics and viceversa.
As we said before U-Statistics have been introduced independently by different people. In particular Halmos was interested in the theory of unbiased estimation, noting that in the case of i.i.d. observations $X_1, \cdots, X_n$

$$E(U_n(h)) = Eh(X_1, X_2)$$

Hence $U_n(h)$ is unbiased estimator of the functional $\theta = \theta(F) = E_F h(X_1, X_2)$, where E_F indicates that each random variables $\{X_i\}$ have the same marginal distribution F. Hoeffding stressed the fact that U-Statistics are a generalized mean, namely of the terms $h(X_i, X_j), 1 \leq i < j \leq n$, and that one could still show asymptotic normality as in the case of ordinary means. Von Mises-Statistics originated in the theory of differentiable statistical functionals, initiated by von-Mises. We will focus on U-Statistics and many sample statistics can be expressed at least approximately as U-Statistics or von-Mises-Statistics, thus providing a very practical reason for the study of these classes of statistics. Below we list some examples from standard textbook examples to some recent applications in the area of dimension estimation of distribution with a fractal support.

SAMPLE VARIANCE

The sample variance is defined as

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

where $\overline{x}$ is the sample mean. Some small calculations show that s_x^2 is a U-Statistic with

kernel $h(x,y) = \frac{1}{2}(x-y)^2$. In fact

$$U_n(h) = \frac{2}{n(n-1)} \sum h(x,y) = \frac{2}{n(n-1)} \frac{n}{2} \sum_{i=1}^{n} (x_i - \overline{x})^2 \tag{4.7}$$

GRASSBERGER-PROCACCIA ESTIMATOR

One of the recent examples of U-Statistics concerns the estimation of fractal dimensions. One such notion is the correlation integral, associated to distributions F on $\mathbb{R}^m$. In general we do not know the distribution F but we have only a finite sample $x_1, \ldots, x_n$ of observations from a stationary process. In the chapter 2 we can find the correlation integral definition in (2.7) and a natural estimator for $C(\epsilon)$ is the sample analogue

$$\begin{aligned} C_n(\epsilon) &= \frac{1}{\binom{n}{2}} \#\{1 \leq i < j \leq n : |x_i - x_j| \leq \epsilon\} \\ &= \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} 1_{\{|x_i - x_j| \leq \epsilon\}} = \hat{C}(\epsilon) \end{aligned} \tag{4.8}$$

which is a U-Statistic with kernel $h(x,y) = 1_{\{|x-y| \leq \epsilon\}}$.

As we said before the Grassberger-Procaccia correlation dimension is based on the kernel function as indicator function $h(x,y;\epsilon) = \mathbf{1}_{\{|x-y| \leq \epsilon\}}$ while gaussian correlation integral is based on $h(x,y;\epsilon) = e^{-\|x-y\|^2/4\epsilon^2}$.The difference between kernel functions consists in the form of kernel it is clear to understand that gaussian kernel is smoother than hard one (see chapter 2); in this view also the Gaussian Correlation Integral is a U-Statistics.

TAKENS ESTIMATOR

An alternative estimator to log-log procedure for the correlation dimension is the Takens estimator. Under the assumption showed in the chapter 2 we obtain

$$\hat{\alpha}_T = -\left(\frac{2}{n(n-1)} \sum_{i<j} \log \frac{|\mathbf{x_i} - \mathbf{x_j}|}{\epsilon_0} \right)^{-1} \tag{4.9}$$

If we replace the independent copies $D_1, \cdots, D_n$ by those dependent pair distances $|x_i - x_j|,\ 1 \leq i,\ j \leq n$, that satisfy $||x_i - x_j|| \leq \epsilon_0$, we obtain the Takens estimator with kernel function as:

$$\log \frac{|\mathbf{x_i} - \mathbf{x_j}|}{\epsilon_0} \tag{4.10}$$

4.2 Proprieties and theorems

Here we will show proprieties and theorems related to U-Statistics. Strong consistency of the U-Statistics of general stationary and ergodic sequences is established by the following theorem of Aaronson et al.

Theorem 3 (Consistency - Aaronson et al. 1). *Let $(X_n)_{n\in N}$ be a stationary ergodic process with a marginal F, and let $h : R^m \longrightarrow R$ be measurable bounded by an F-continuous function. Then*

$$\lim_{n\longrightarrow\infty} U_h(X_1, \cdots, X_n) = \theta_h(F) \ \ a.s$$

In the case of absolutely regular sequences, a result similar to Theorem 3 was established under milder conditions on the kernel function no continuity condition is required in this case.

Theorem 4 (Consistency - Aaronson et al. 2). *Let $(X_n)_{n\in N}$ be a stationary absolutely regular process with a marginal F, and let $h : R^m \longrightarrow R$ be measurable bounded function. Then*

$$\lim_{n\longrightarrow\infty} U_h(X_1, \cdots, X_n) = \theta_h(F) \ \ a.s$$

One of the main tools in studying U-Statistics is the Hoeffding's decomposition also called projection method. It says that every U-Statistic can be written as a finite weighted sum of degenerate U-Statistics. A U-Statistic of degree 2 can be decomposed

as

$$U_n(h) = \theta(F) + \frac{2}{n}\sum_{i=1}^{n}[h_1(X_i - \theta(F))] + R_n \tag{4.11}$$

where

$$h_1(x) = \int h(x,y)dF(y) \tag{4.12}$$

for more precise comment see later.

The term R_n called remainder, is actually defined by the relation (4.11) in this case it can be written as

$$R_n = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}[h(X_i,X_j) - h_1(X_i) - h_1(X_j) + \theta(F)] \tag{4.13}$$

If $U_n(h)$ itself is not a degenerate, R_n is small comparing to the leading term $\frac{2}{n}\sum_{i=1}^{n}[h_1(X_i - \theta(F))$, which dominates the asymptotic behaviour and the variance of the U-Statistics.

To prove the the central limit theorem (CLT) for U-Statistics we can see two different theorems, here we propose the Denker-Keller version[1] because they consider some additional conditions on the functional and the kernel h. This extension is of particular importance for data from dynamical systems.

Theorem 5 (Denker-Keller). *Let $h : R^{k\times k} \longrightarrow R$ be a non degenerate kernel. Then the asymptotic distribution of $\frac{\sqrt{n}}{m\sigma_n}(U_n(h) - \theta)$ is $N(0,1)$ provided one of the following conditions in satisfied:*

1. *$(X_n)_{n\geq 1}$ is uniformly mixing in both directions of time, $\sigma_N^2 \longrightarrow \infty$ and for some $\delta > 0$*

$$\sup E|h(X_{t1},\cdots,X_{tm})|^{2+\delta} < \infty$$

2. *$(X_n)_{n\geq 1}$ is uniformly mixing in both directions of time with mixing coefficient $\phi(n)$*

[1] In independent way also Yoshihara [1976]. They found the same results using different tools.

satisfying $\sum \phi(n) < \infty$, $\sigma^2 \neq 0$ *and*

$$\sup E|h(X_{t1}, \cdots, X_{tm})| < \infty$$

3. $(X_n)_{n\geq 1}$ *is absolutely regular with coefficients* $\beta(n)$ *satisfying* $\sum \beta(n)^{\delta/2+\delta}$ *for some* $\delta > 0$, $\sigma^2 \neq 0$ *and*

$$\sup E|h(X_{t1}, \cdots, X_{tm})|^{2+\delta} < \infty$$

□

In this particular case we have that n is the number of observation in the time series, m number of kernel elements equal to $m = 2$ and so we obtain:

$$\sqrt{n}(U_n(h) - \theta) \sim N(0, 4\sigma^2)$$

and

$$\sqrt{Var(U_n(h))} = \frac{2\sigma}{\sqrt{n}}$$

The asymptotic variance of a U-Statistics is given by the Hoeffding decomposition

$$\sigma^2 = \mathrm{Var}[h_1(X)] + 2\sum_{k=1}^{\infty} \mathrm{Cov}[h_1(X_i), h_1(X_{k+i})] \tag{4.14}$$

It is needed to define $h_1(x)$ that is a particular kernel function which is given by:

$$h_1(x) = \int h(x, y)dF(y),$$

called degenerative U-Statistics. If we consider the hard kernel $h(x, y) = \mathbf{1}_{\{|x-y|leq\epsilon\}}$ then $h_1(x)$ is:

$$h_1(x) = \int \mathbf{1}_{\{|x-y|\leq\epsilon\}} dxF(y) = P_y(|x - y| \leq \epsilon)$$

that is the mass of ball of radius ϵ around X. Given a sample $x_1, x_2, \ldots, x_n$, the sample

quantity for $h_1(x)$ will be:

$$\hat{h}_1(x) = \frac{1}{n-1} \sum_{i=1}^{n} \mathbf{1}_{\{|x-x_i| \leq \epsilon\}}$$

When we use other kernel we loose clear meaning but the most important role of $h_1(x)$ is to simplify variance computation using Hoeffding decomposition.

So we have that:

$$h_1(x) = \int h(x,y) dF(y) = \int \exp\left(\frac{-\|x-y\|^2}{4\epsilon^2}\right) dF(y)$$

using the Gaussian kernel we loose the precise meaning. The sample quantity to investigate is:

$$4\hat{\sigma}^2 = \hat{\sigma}_U^2 = 4\text{Var}(\hat{h}_1(x) + 8 \sum_{k=1}^{M} \text{Cov}[\hat{h}_1(x_i)\hat{h}_1(x_{k+i})] < \infty \tag{4.15}$$

where M is the lag of covariance function after that the function become flat.

Chapter 5

Results - Theoretic aspects

In this chapter we will show some asymptotic proprieties of Gaussian Correlation Integrals and Takens Dimension. To do it we will use the conditions of Denker-Keller theorem (th.5). We can use theorem because our quantities are kernel functions. So a direct consequence of U-Statistics theorems are the proofs of normality distribution of our quantities. Another characteristic to investigate is the consistency which is straight to compute. So in this chapter we will start to show the consistency and later normality. In the part we will show some practical results.

5.1 Gaussian Correlation Integral - Consistency and Normality

In this section we will prove consistency and normality of Gaussian Correlation Integral.

Strong consistency of the U-Statistics of general stationary and ergodic sequences is established by the theorem of Aaronson et al. showed in the last chapter (th. 3) In that way to prove the consistency we use the theorem conditions. It is straight to see that under the conditions of the time series being stationary and ergodic the Gaussian correlation integral is consistent because the kernel function is everywhere bounded and continuous. Borovkova et al. [2001]showed the same property for the hard kernel and for the kernel in the Takens estimator.

An interesting role here is taken by the asymptotic normal behavior of invariants.

If the consistency was investigated using Aaronson theorems as a prompt consequence, this was not possible for normality. Actually the normality is also a very straightforward application of Denker-Keller theorem (th. 5). We would not formulate our results as separate theorem. Considering the second condition of the theorem we

can show the normal distribution of the U-Statistics and in particular of the Gaussian Correlation Integral $T_m(\epsilon)$.

Corollary 1 (Normality of Gaussian Correlation Integral $T_m(\epsilon)$). *Let $h : R^2 \longrightarrow R$ be a non-degenerate kernel. Then the asymptotic distribution of $\frac{\sqrt{n}}{2\sigma_n}(U_n^T - \theta)$ is $N(0,1)$ if $(X_n)_{n\geq 1}$ is absolutely regular $\sigma^2 \neq 0$ and*

$$\sup E|h(X_{t1}, \cdots, X_{tm})|^{2+\delta} < \infty$$

The second condition focus on the asymptotic variance and on the bound of sup of the kernel function expected value. In fact if $(X_n)_{n\geq 1}$ is uniformly mixing in both directions of time with mixing coefficient $\phi(n)$ satisfying $\sum \phi(n) < \infty$, $\sigma^2 \neq 0$ and

$$\sup E|h(X_{t1}, \cdots, X_{tm})| < \infty$$

then the Denker-Keller theorem hold.

PROOF Assuming that the process is uniformly mixing in both direction of time with mixing coefficient $\phi(n)$, we know that

$$\sup E(h(x_i, x_j)) \quad = \sup E\left(\exp \frac{-|x_i - x_j|^2}{4\epsilon^2}\right) \tag{5.1}$$

and by the definition of expected value we obtain

$$E\left(\exp \frac{-|x_i - x_j|^2}{4\epsilon^2}\right) \quad = \int \exp \frac{-|x_i - x_j|^2}{4\epsilon^2} f(x_i) f(x_j) dx_i dx_j \tag{5.2}$$

The function on the right side inner the sum, i.e. $\exp \frac{-|x_i - x_j|^2}{4\epsilon^2}$, has the same form of $\exp(-t)$ where $t > 0$ and ϵ constant. So we have that

$$0 < E\left(\exp \frac{-|x_i - x_j|^2}{4\epsilon^2}\right) < 1$$

and it is finite and of course its sup. □

Here we want to give the distribution of $\log T_m(\epsilon)$ based on the *U-Statistics*. We have already known that $T_m(\epsilon)$ is Normal distributed so we have to find new parameters

for the distribution; to do it it is needed to apply Serfling theorem.

Theorem 6 (Serfling - univariate case). *Suppose that X_n is $AN(\mu, \sigma_n^2)$, with $\sigma_n \to 0$, where σ_n^2 means the variance depends on n. Let g be a real-valued function differentiable at $x = \mu$, with $g(\mu) \neq 0$. Then*

$$g(X_n) \text{ is } AN(g(\mu), [g`(\mu)]^2\sigma_n^2)$$

□

From Denker-Keller theorem, we know:

$$\sqrt{n}(T_m(\epsilon) - \theta) \sim AN(0, 4\sigma^2) \tag{5.3}$$

where θ is

$$\theta = \int\int h(x_i, x_j)dF(x_i)dF(x_j)$$

$$T_m(\epsilon) \sim AN(\theta, 4\sigma^2/n) \tag{5.4}$$

considering the g-function is $\log(x)$ so applying the theorem the asymptotic distribution of $\log T_m(\epsilon)$ became:

$$\log T_m(\epsilon) \sim AN\Big(\log(\theta), \frac{4\sigma^2}{n\theta^2}\Big) \tag{5.5}$$

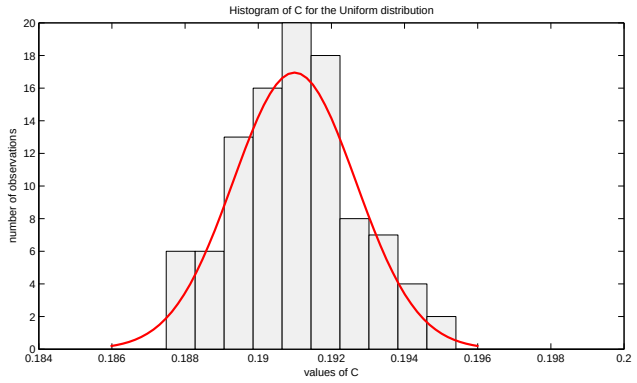
and θ and σ should be estimate.

In particular θ is estimated by the sample mean and σ is estimated using the expression of variance via $h_1(x_i) = \int \exp(\frac{|x_i - x_j|^2}{4\epsilon^2})dF(x_j)$. We will test in the next section the role of h_1-variance respect to the computation variance using the definition on the basis of generated data using Hènon map, Lorenz system and independent data.

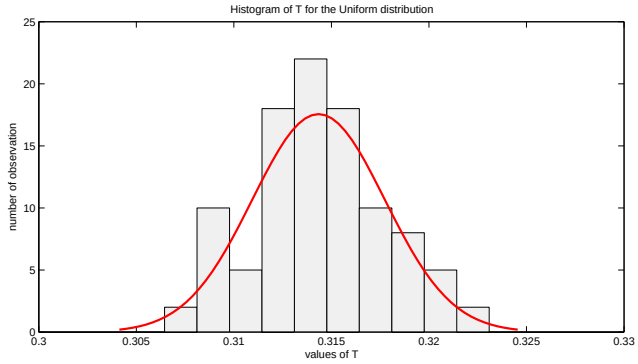
5.1.1 Application

An example of this distribution is given for the Uniform case. This is one of the most popular example, in fact a lot of studies show how the Uniform distribution works. We

generate a set $B = 100$ of Uniform distributions in the interval $(0, 1)$ of length 1000 and over each vector of 1000 points we compute the Gaussian Correlation Integral. So we obtain a vector carried out 100 $T(\epsilon)$ value. In the Figure 5.1., we can see the behavior.



(a)



(b)

Figure 5.1. Normal Distribution for (a) $C(\epsilon)$ and (b) $T(\epsilon)$ in Uniform distribution case

5.2 Takens estimator - Normality

As said before, the consistency was proved by Borovkova [1998], so here we investigate the Normality. Another application of the Denker-Keller theorem is related to the Takens estimator for the Correlation Dimension.

Again the Takens estimator is an example of U-Statistics[1] so we can use the theorem

[1]We have consider it as an example of U-Statistics in the chapter 3

to show the asymptotic distribution of U_n^T and $\hat{\hat{\alpha}}_n^T$, where

$$U_n^T = \log \frac{|x_i - x_j|}{\epsilon_0}$$

and

$$\hat{\alpha}_n^T = -\left(\frac{2}{n(n-1)} \sum_{i<j} \log \frac{|\mathbf{x_i} - \mathbf{x_j}|}{\epsilon_0} \right)^{-1}$$

Corollary 2 (Normality of Takens estimator). *Let $h : R^2 \longrightarrow R$ be a non-degenerate kernel. Then the asymptotic distribution of $\frac{\sqrt{n}}{2\sigma_n}(U_n^T - \theta)$ is $N(0,1)$ if $(X_n)_{n\geq 1}$ is absolutely regular $\sigma^2 \neq 0$ and*

$$\sup E|h(X_{t1}, \cdots, X_{tm})|^{2+\delta} < \infty$$

PROOF If the process is absolutely regular then it is possible to use the kernel function $h(x_i, x_j)$ even if the kernel is not a bounded function. We have to investigate the condition on the $2+\delta$ moment (as for $T(\epsilon)$):

$$\sup E \left| \log \frac{|x_i - x_j|}{\epsilon_0} \right|^{2+\delta} < \infty$$

Let $X_i \in A$ for all i, where A is some bounded subset of R^k. Note that, if the joint density function of (x_i, x_j) has a density f_{ij} which is bounded by C, then:

$$\begin{aligned} E\left|(\log|x_i - x_j|)\right|^{2+\delta} &= \int\int \left|(\log|x_i - x_j|)\right|^{2+\delta} f_{ij} dx_i dx_j \\ &\leq C \int_{A^2} \int \left|(\log|x_i - x_j|)\right|^{2+\delta} dx_i dx_j < \infty \end{aligned}$$

It is true because of integration over a bounded set (A is a bounded set). If we consider that $\rho_{ij} = |x_i - x_j|$ has density $p(x)$ we can show that $\int |\log r|^{2+\delta} p(r) dr$ is finite for every $p(x)$ such that $p(x) \sim O(x^{-\delta})$ as $x \to 0$ because we can divide the integral in two parts and so we'll have:

$$\int |\log r|^{2+\delta} \frac{1}{r^\delta} dr = \int_0^\epsilon (\log r)^{2+\delta} \frac{1}{r^\delta} dr + \int_\epsilon^{r_0} (\log r)^{2+\delta} \frac{1}{r^\delta} dr \tag{5.6}$$

In $[0, \epsilon]$ it is known $\log x \sim x$ so the integral is defined and finite. □

After this theorem we know that:

$$U_n^T \sim N(\theta, 4\sigma^2/n) \tag{5.7}$$

where $\sigma^2 = Var[h_1(X)] + 2\sum_{k=1}^{\infty} Cov[h_1(X_1), h_1(X_{k+1})]$.

It is important to define

$$h_1(x) = \int h(x,y) dF(y) = \int \log\Big(\frac{|x-y|}{r_0}\Big) dF(y)$$

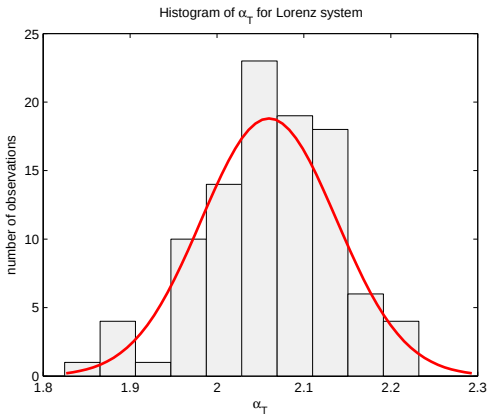
It is important to remember that the Takens estimator is $-\frac{1}{U_T}$ so it is necessary to prove that this random variable is normally distributed. To have the right distribution of α_n^T we have to apply the Serfling theorem.

In our case we have prove that $U_n^T \sim AN(\theta, 4\sigma^2/n)$. The transformation function used to obtain d is $g(x) = -\frac{1}{x}$ and it is differentiable for every x such that $x \neq 0$. Now this theorem can be applied to Takens' estimator and we'll obtain:

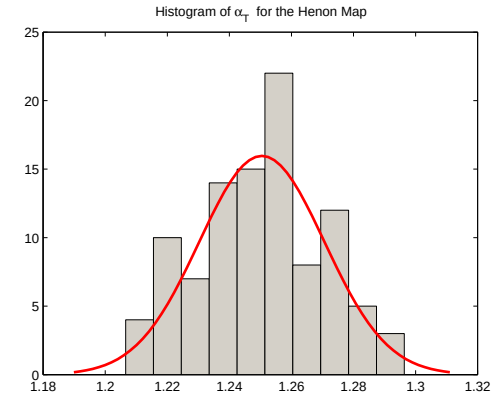
$$\alpha_n^T \sim AN\Big(-\frac{1}{\theta}, \frac{4\sigma^2}{n\theta^4}\Big)$$

5.2.1 Application

Here we propose as an example applications related to the Lorenz system and the Hènon map. Choosing the upper cut-off it is possible to find an estimate in agreement to the literature for $\epsilon_0 = 23.25$ for Lorenz and $\epsilon_0 = 2.1$ for Hènon map. The histograms show the behavior of the Takens estimator. The distributions satisfy the Lilliefors-test and in Figure 5.2 it is possible to see the behaviors of the distribution.



(a)



(b)

Figure 5.2. Histogram of $\hat{\alpha}_n^T$ fixed $\epsilon_0 = 23.25$ and $m = 8$ in the Lorenz system (a) and histogram of $\hat{\alpha}_n^T$ fixed $\epsilon_0 = 2.1$ and $m = 2$ in the Henon Map (b)

Chapter 6

Results - Practical aspects

In this section, practical results of our theoretic developments will be shown. In particular we consider well known systems in literature as Henon map, Lorenz system and Uniform distribution to prove the theoretic results.

Using U-Statistics method, we are able to compute on one single time series an accurate estimate both for the Correlation Integral $C_m(\epsilon)$ and for the Gaussian Correlation Integral $T_m(\epsilon)$ and a reliable estimate of the standard deviation for its.

Results using U-Statistics tools will be compared to real data which represent the benchmark value. The work is organized in the following way:

1. empirical proof of the Normal behavior for the $T(\epsilon)$. Takens normality can be seen in Figure 5.2;

2. comparison with value related to Gaussian Correlation Integral values and its variance using analytical method and the real data;

3. comparison between the distributions of the correlation dimension computed using analytical method and real data.

The analysis starts choosing ϵ_i values then computing the $C_m(\epsilon_i)$ and $T_m(\epsilon_i)$. Invariants values will be showed for some particular (in this case only three) ϵ_i values taken in the interval of $5 - 10\%$ of the length of the system. For each ϵ_i we create a number of B replications of $C_m(\epsilon_i)$ and $T_m(\epsilon_i)$. Empirical distribution of the Correlation and Gaussian Correlation Integral, fixed m and ϵ, are obtained. It is possible to investigate the agreement to the Normal distribution. We can see the behavior plotting

the histogram and its normal fit then the validity is given by testing the null hypothesis of Normal distribution as **Corollary 1** suggests. Our results show that for all ϵ values, we accept the hypothesis that $C_m(\epsilon_i)$ and $T_m(\epsilon_i)$ are Normal distributed. This consideration is supported by the Lilliefors test.

Lilliefors test is a single sample Lilliefors hypothesis test of composite normality. It performs the Lilliefors modification of the Kolmogorov-Smirnov test to determine, if the null hypothesis of composite normality is a reasonable, assumption regarding the population distribution of a random sample X (in this case $C_m(\epsilon_i)$ or $T_m(\epsilon_i)$).

The Lilliefors test is a 2-sided test of composite normality with sample mean and sample variance used as estimates of the population mean and variance, respectively. The test statistic is based on the 'normalized' samples

We start the analysis considering free time series, then we add measurement noise normal distributed with mean equal to 0 and different standard deviation values for each system considered. The presence of noise should be smaller to avoid abrupt estimates. In this work we consider well known systems described in the following sections.

6.1 Independent case

In this part we consider the independent situation. It is shown as an example. We have already investigated the system in the last chapter so here we propose the agreement with literature. Let continuous Uniform distribution be in $[0, 1]$ interval, we create independent time series from the distribution of length 1000. For several ϵ values we compute the Gaussian Correlation Integral. In this situation we consider as region scaling following values: $\epsilon = (0.1, 0.075, 0.05)$.

Values given in the table 6.1 are similar to values present in Theiler [1990]. In fact using the ordinary metric[1] the value for $\epsilon = 0.1$ is equal to $C(0.1) \simeq 0.19$ as in our computation, so we are able to give values using analytical method related to only one time series.

We can observe normal distribution of the Gaussian Correlation Integral as theory suggests. In the last chapter, we present an example of the Normal distribution behavior see Figure 5.1 for a particular single value $\epsilon = 0.1$ both Gaussian and Correlation Integral.

[1] Euclidean metric

ϵ	$\mathbf{C_A}(\epsilon)$	$\mathbf{sd_A^C}$	$\mathbf{T_A}(\epsilon)$	$\mathbf{sd_A^T}$
0.05	0.0982	$7.7 \cdot 10^{-4}$	0.1660	0.0015
0.075	0.1443	0.0012	0.2421	0.0025
0.1	0.1891	0.0018	0.3135	0.0035

Table 6.1. Values of C and T and its standard deviation using the analytical method based on U-Statistics for Uniform distribution.

6.2 Henon map

The Henon map is one of the most famous chaotic equation based on a particular system of equations as follows:

$$\begin{aligned} x_{t+1} &= y_{n+1} - ax_t^2 \\ y_{t+1} &= bx_t \end{aligned} \tag{6.1}$$

The map depends on two parameters a and b which for values of $a = 1.4$ and $b = 0.3$ the map behavior became chaotic. In the next figure we can see the Henon map and then, using the suitable embedding parameters $\tau = 1$ and $m = 2$ we can obtain the embedded map knowing only the first component.

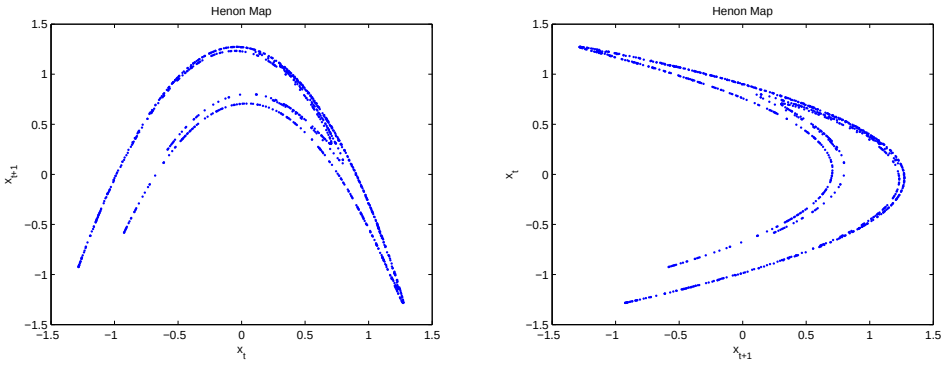
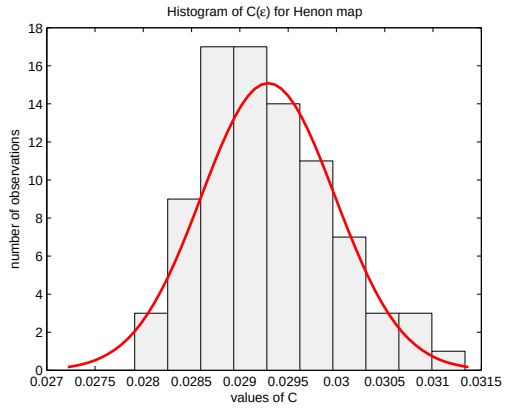


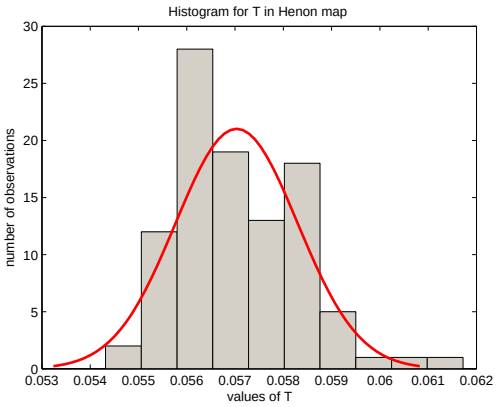
Figure 6.1. Simple and Embedded Henon Map

Here we can see the agreement to the **Corollary** 1, so we can show the normal distribution of the Gaussian and the Correlation Integral using $k = 100$ true time series from the Hènon map; in Figure 6.2 we can see the behaviors.

In the table 6.2 we will show how is the agreement degree between the analytical and real computation in several cases. In particular the analytical result is obtained using the formula in (2.31); the computation of the $T(\epsilon)$ values and the standard deviation



(a)



(b)

Figure 6.2. Henon Map. Normal Distribution for C (a) and T (b) for a fixed $m = 2$ and for a fixed $\epsilon = 0.1$ without noise.

is computed using (4.15). We choose $M = 50$ value for stopping covariance term; after that lag the sum of covariance remains more or less constant. The main advantage of the formula is that we are able to compute straight, and at the same time, the value of the invariant and its standard deviation. Otherwise, using real data we are able to compute over 100 time series the 100 values for the invariant but we obtain only one estimated value of the variance, so we have needed to do this procedure several times for obtaining the sample of the variance. In this work we indicate with "A" results obtained by analytical approach and with "R" results by real data.

In Chapter 2 we showed the log-log plot then we had to find the estimate of the correlation dimension as the slope of the line in the linear zone. It is possible to obtain the log-log plot both Correlation and Gaussian Integral. The estimate of the correlation

ϵ	$\mathbf{C_R}(\epsilon)$	$\mathbf{C_A}(\epsilon)$	$\mathbf{sd_R^C}$	$\mathbf{sd_A^C}$
0.05	0.0198	0.0134	$5.8 \cdot 10^{-4}$	$6.05 \cdot 10^{-4}$
0.075	0.0316	0.0216	$7.1 \cdot 10^{-4}$	$8.42 \cdot 10^{-4}$
0.1	0.0447	0.0306	$9.0 \cdot 10^{-4}$	$1.00 \cdot 10^{-3}$
	$\mathbf{T_R}(\epsilon)$	$\mathbf{T_A}(\epsilon)$	$\mathbf{sd_R^T}$	$\mathbf{sd_A^T}$
0.05	0.0398	0.0260	$8.20 \cdot 10^{-4}$	$7.80 \cdot 10^{-4}$
0.075	0.0645	0.0420	$1.40 \cdot 10^{-3}$	$1.13 \cdot 10^{-4}$
0.1	0.0916	0.0591	$2.00 \cdot 10^{-3}$	$1.40 \cdot 10^{-3}$

Table 6.2. Henon map case. Values of $C(\epsilon)$ and $T(\epsilon)$ using the analytic tools based on U-Statistic and comparison with real data.

dimension using Gaussian Correlation Integral is $\hat{\alpha} = 1.18$ in $(0.02 - 1)$ which is close to the known value 1.2.

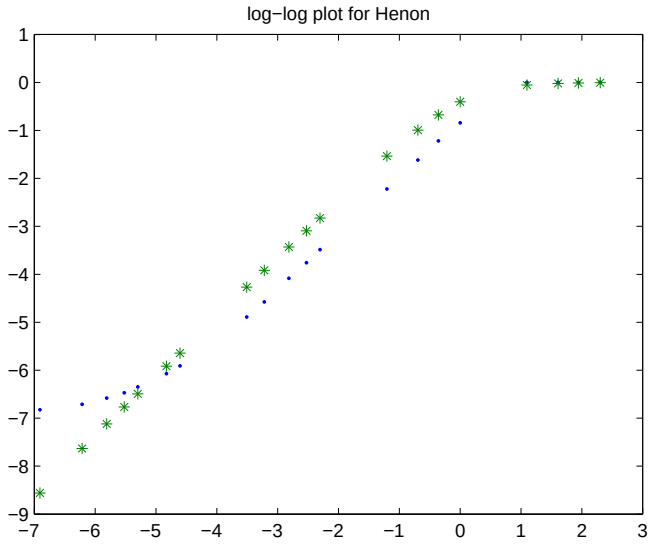


Figure 6.3. Comparison for the log-log using the Correlation Integral $C(\epsilon)$ (dot line) and the Gaussian Correlation Integral $T(\epsilon)$ (star line) in the Henon map.

We can obtain the value of $T(\epsilon)$ and its s.d. by using one time series. In the next step we will found the correlation dimension distribution from analytical method. In fact, we know the scaling region over which the right correlation dimension is found. We know how is the asymptotic distribution of the Gaussian Correlation Integral (5.4), so we can generate $B = 1000$ log-log lines and over each line the correlation dimension is estimated. Then, the correlation dimension histogram based on the analytical way is

	α	sd
analytic	1.2089	0.0322
real	1.2087	0.0336

Table 6.3. Values of the distribution of the correlation dimension using the analytic method in free case for Henon map.

obtained. The agreement to the reality is proved comparing the analytical method to the results from real data. It is possible to look at the agreement in the Figure 6.4.

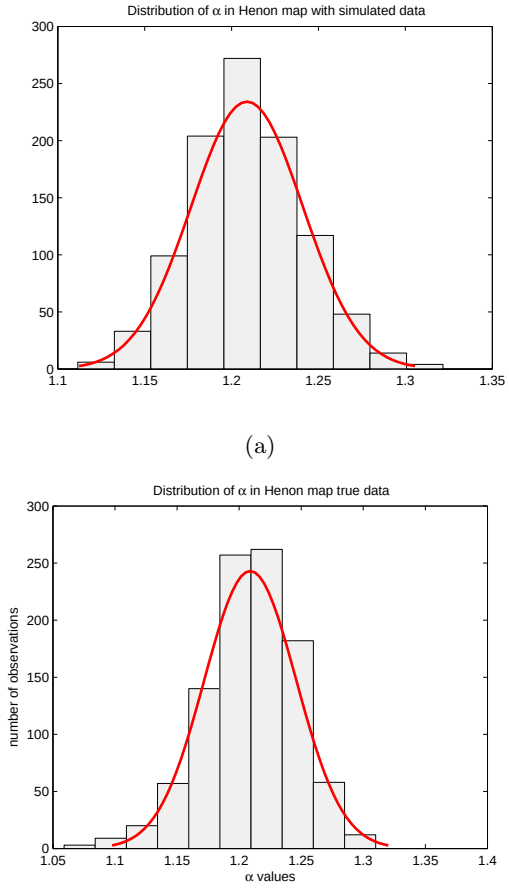


Figure 6.4. Henon map. Histogram of the correlation dimension α using analytical method (a) and real data (b).

In the table 6.3 we can observe the good agreement related to the estimate of location (mean) and scale (s.d.) parameter of distributions; analytical and real approaches produce the same conclusions.

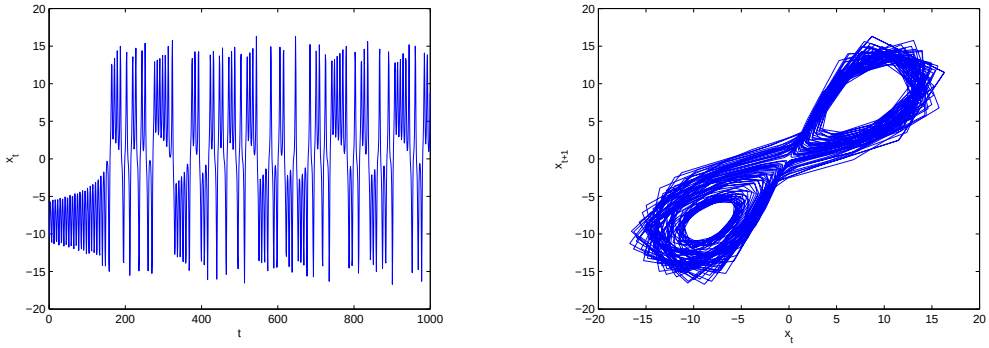
It is important to see how the correlation dimension estimates in the Henon map are distributed in a range between 1.1 and 1.3, so the null hypothesis of fractal dimension is accepted. This conclusion is not always obvious: an example of this situation is given by the Lorenz system.

6.3 Lorenz system

Here we will make the same analysis using the Lorenz system. The attractor itself and the equations are derived from the simplified equations of convection rolls arising in the equations of the atmosphere. From a technical standpoint, the system is nonlinear, three-dimensional and deterministic. The equations governing the Lorenz attractor are:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z \end{aligned} \tag{6.2}$$

Usual chaotic values for parameters are $\sigma = 10$, $\beta = 8/3$ and $\rho = 28$. We consider the embedding parameters given in literature i.e. $\tau = 8$ and $m = 3$ and we apply these parameters to the x-component of the system. In Figure 6.5 we show the component useful for our analysis.

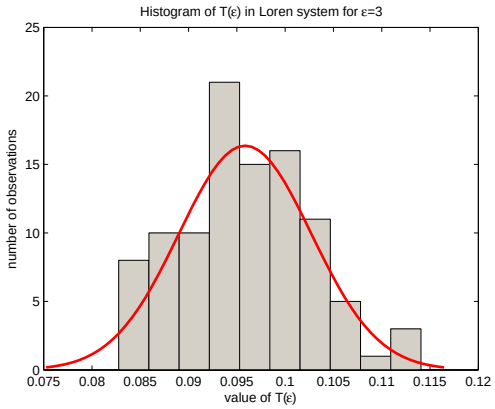


(a) (b)

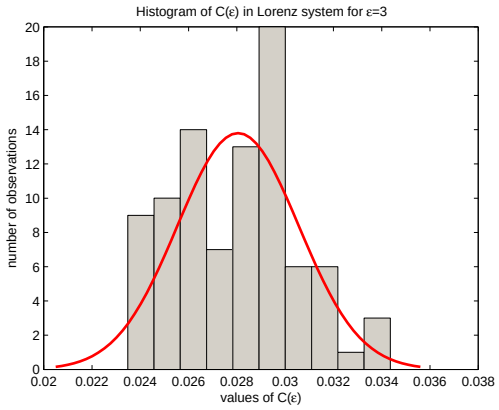
Figure 6.5. x-component of the Lorenz system (a) used to obtain the embedding phase space and the x-component plotted in the scatter-plot (b)

In the continuous space, Lorenz system is one of most popular example of dynamic

system. Here we can see the agreement to the **Corollary** 1, so we can show the normal distribution of the Gaussian and the Correlation Integral using $k = 100$ true time series from the Lorenz system; in Figure 6.6 we can see the behaviors.



(a)



(b)

Figure 6.6. Lorenz system. Normal Distribution for C (a) and T (b) for a fixed $m = 3$ and for a fixed $\epsilon = 3$.

As in Hènon map we can obtain the agreement to the invariants of the system comparing the outcomes from analytical method to the outcomes based on the real data. We choose $M = 15$ value for stopping covariance term; in fact after this lag, the sum of covariance remains more or less constant.

In the table 6.4 we show results related to these approaches. In particular we can see how the estimated values for the $C(\epsilon)$, $T(\epsilon)$ and their standard deviations are in agreement in both cases. Here we consider Lorenz system of length $n = 1000$ and with the sample rate 0.01 without measurement noise.

ϵ	$\mathbf{C_R}(\epsilon)$	$\mathbf{C_L}(\epsilon)$	$\mathbf{sd_R}$	$\mathbf{sd_L}$
1.5	0.0050	0.0074	$4.90 \cdot 10^{-4}$	$2.96 \cdot 10^{-4}$
2.25	0.0138	0.0163	$9.90 \cdot 10^{-4}$	$8.51 \cdot 10^{-4}$
3	0.0246	0.0297	$1.63 \cdot 10^{-3}$	$1.80 \cdot 10^{-3}$
	$\mathbf{T_R}(\epsilon)$	$\mathbf{T_L}(\epsilon)$	$\mathbf{sd_R}$	$\mathbf{sd_L}$
1.5	0.0245	0.0284	0.0015	0.0015
2.25	0.0523	0.0609	0.0026	0.0025
3	0.0862	0.0991	0.0034	0.0029

Table 6.4. Lorenz system case. Values of $C(\epsilon)$ and $T(\epsilon)$ using the analytic tools based on U-Statistic and comparison with real data.

The estimate of the correlation dimension is given by a log-log plot Figure 6.7 in the linear zone. It is possible to find the estimate using both Gaussian and Correlation Integral. It is needed to look the linear zone then the correlation dimension is found as the slope of the line. As for Hènon, the estimated correlation dimension obtained using the log-log plot is in agreement to the literature; our result is $\alpha = 2.067$ in the linear zone $(0.08 - 25)$ and the real value is around 2.06.

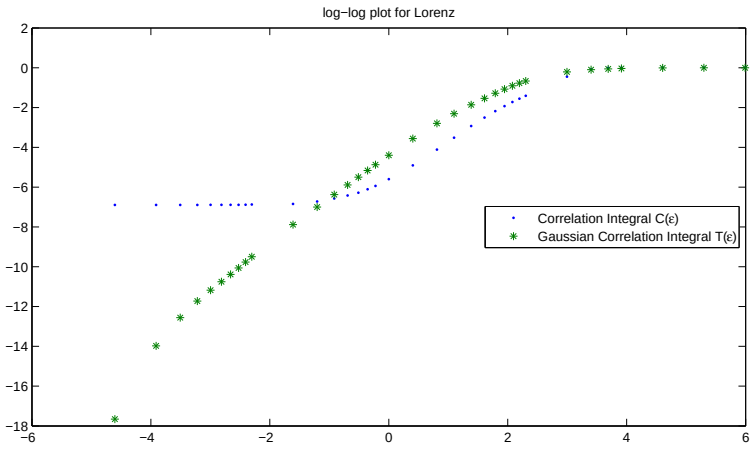
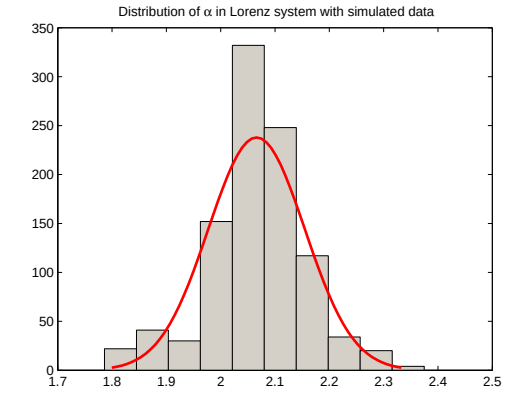


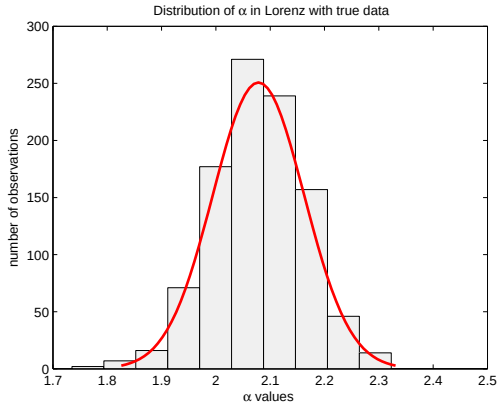
Figure 6.7. Comparison for the log-log using the Correlation Integral $C(\epsilon)$ and the Gaussian Correlation Integral $T(\epsilon)$.

We can find the distribution of the correlation dimension in the Lorenz system using analytical and real approaches.

In the table 6.3 we can observe the good agreement between different approaches in terms of values computed, in fact mean and standard deviation are close.



(a)



(b)

Figure 6.8. Histogram of the correlation dimension using analytical method (a) and real data (b) in Lorenz system.

	α	sd
analytical	2.066	0.088
real	2.078	0.084

Table 6.5. Values of the distribution of the correlation dimension using the analytic method in free case for Lorenz system.

We can compare both approaches looking histograms. Graphs show an equal proportion of 20% of α values that are less than 2: this means that we are not able to assure that the system has a fractal dimension even if we know that the Lorenz system is fractal. This conclusion represents a problem in the correlation dimension estimate so it is needed to link the estimate to its standard deviation. The rejection of the hypothesis of fractal dimension, even if we know that the system is fractal, is caused

by a several factors. One of them is surely the time series length: each system and so each time series have its length. In the real case we have short time series, so it would be useful to check a practical method that could be efficient respect to the time series length.

6.4 Noise Presence

In this section we want to show a first trial of the method proposed in noise presence. In fact this situation is more common. In the real world, phenomena are the result of a pure signal and a noise presence. Here, we add low noise level otherwise the system could be change its equilibrium. We consider Hènon map and Lorenz system as in above.

6.4.1 Noise presence - Hènon

Here we will investigate the Hènon map. We add to the signal a normal distributed noise with mean equal to 0 and standard deviation equal to 1.

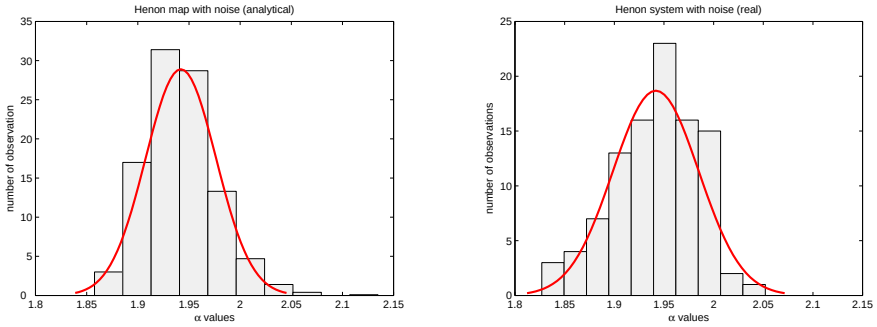
Considering, as in free time series, three values of $\epsilon = (0.1, 0.075, 0.05)$ so we obtain the table 6.6 of values $T^n(\epsilon)$ and $C^n(\epsilon)$ and its standard deviation, where n represents noise presence.

ϵ	$T^n(\epsilon)$	sd_{T^n}	$C^n(\epsilon)$	sd_{C^n}
0.05	$1.6 \cdot 10^{-3}$	$1 \cdot 10^{-4}$	$1.4 \cdot 10^{-3}$	$5.3 \cdot 10^{-5}$
0.075	$3.5 \cdot 10^{-3}$	$2 \cdot 10^{-4}$	$1.9 \cdot 10^{-3}$	$9.0 \cdot 10^{-5}$
0.1	$6.3 \cdot 10^{-3}$	$4 \cdot 10^{-4}$	$2.6 \cdot 10^{-3}$	$1.3 \cdot 10^{-4}$

Table 6.6. Hènon Map. Noise presence. Values of $C^n(\epsilon)$ and $T^n(\epsilon)$ computed using analytical way and its standard deviation.

In the interval $(0.01 - 0.7)$, it is possible to see the linear zone so the correlation dimension estimate in this interval is equal to $\alpha = 1.94$. Following the same procedure used in absence of noise we compute for a set of ϵ (the same used in free noise case) values of Gaussian Correlation Integral and its standard deviation. We obtain the correlation dimension distribution Figure 6.9.

When we used the analytical method the distribution is related to $B = 1000$ valued of $\hat{\alpha}$ while using real data we take in consideration just considered $B = 100$ values. The stopping term used in the covariance sum was $M = 50$. The two distribution are Normal distributed with mean and standard deviation summarized in the table 6.7.



(a) (b)

Figure 6.9. Distribution of α in Hènon map in noise presence computed using the analytical method (a) and real data (a).

There is agreement in the values, so the procedure is good, but it is clear that the α estimate goes far from the real value even if low noise presence.

	α^n	sd
analytical	1.94	0.034
real	1.94	0.040

Table 6.7. Values of the distribution of the correlation dimension using the analytic method.

6.4.2 Noise presence - Lorenz

Here we want to investigate how our method works in noise presence time series. We add a normal distributed noise with mean equal to zero and standard deviation equal to 0.05. The noise level added is low because of the system instability. We consider the same set of ϵ values used in free-noise time series. Here we will give table 6.8 of $T^n(\epsilon)$ and $C^n(\epsilon)$ and its standard deviation.

ϵ	$T^n(\epsilon)$	sd_{T^n}	$C^n(\epsilon)$	sd_{C^n}
1.5	0.0285	0.0022	0.0074	0.0006
2.75	0.0611	0.0035	0.0164	0.0015
3	0.0992	0.0039	0.0298	0.0031

Table 6.8. Lorenz system. Noise presence. Values of $C^n(\epsilon)$ and $T^n(\epsilon)$ computed using analytical method and its standard deviation.

In the interval $(0.08 - 25)$ it is possible to find the linear zone and the correlation dimension estimate is 2.051.

The distribution of $B = 1000$ α values both using analytical method and real data is Normal distributed with mean and standard deviation summarized in the table 6.9. There is agreement in the values, so the procedure is good, but it is clear that the α estimate goes far from the real value even if low noise presence. There are almost 20% is less then 2, so

	α^n	sd
analytical	2.10	0.07
real	2.06	0.08

Table 6.9. Lorenz system. Noise presence. Values of the distribution of the correlation dimension using the analytic method and real data.

Chapter 7

Conclusion

The intent of this work is to give an introduction to methods related invariants estimate. Multiple interests of the argument in several fields of Science justify our choice to deal without a surplus of formula and mathematical details. In this view it is possible to address the work also for people unfamiliar in these topics but some people can happen on some particular phenomena. In fact there are many phenomena that look like without a precise structure so it is needed to use chaotic theory for the time series analysis.

The work have given basic elements related to chaos theory and the most used estimators for invariant quantities.

The estimate of the invariants is not simple because it depends on the scaling region. There are some methods and considerations about this problem but none gives a precise solution, such that the estimate process "*is often as much an art as it is a science*". Diks [1999]. In fact, each system has a especially structure so it becomes hard to find an unique rule.

The theoretic innovation of the thesis consists in the checking of some asymptotic proprieties of invariant quantities: consistency and normality of Gaussian Correlation Integral and normality of the Takens estimator. These conclusions result from U-Statistics theorems and they are supported by empirically proof.

Practical innovations concern the computation of the correlation dimension distribution based on the analytical method. It is possible to obtain the distribution using only one time series. Then, we compare distribution comes from using the analytical method to distribution from real data: several considerations can be done. The work underlies that the main statistics problem is the need to give estimate precision, this is

not always simple to obtain. The difficulty is shown by the presence in several studies and work that confirm the fractal system existence without a precision value. For this reason it is important the searching of standard deviation of the estimated quantity. Methods used to compute the standard deviation thanks the analytic way consider observations not independent and it is realistic; on the other side the agreement to the real data is clear so the U-Statistics procedure is realistic.

Problems are related to the inner composition of the system, in fact we can found systems that produce always fractal dimension but sometimes it is not done. In general we are not able to distinguish between fractal or not-fractal set without the confidence interval. An example is given by Lorenz system that is a fractal system, but computations are not able to assert it in free time series. In the noise time series case, the results get worse, in terms of estimation, even if low noise level.

Future developments can be summarized as follows. It would be important to study other well-known systems and to come out agreement between two methods; another important step should be to add noise and to understand what happen; to check the using method with shorter time series.

Bibliography

J. Aaronson, R. Burton, H. Dehiling, D. Gilat, T. Hill, and B. Weiss. Strong laws for L- and U-statistics. *Transactions of the American Matehmatical Society*, 348(7), 1996.

S. Borovkova. *Estimation and prediction for nonlinear time series.* PhD thesis, University of Groningen, 1998.

S. Borovkova, R. Burton, and H. Dehling. Limit theorems for functional of mixing processes with application to U-statistics and dimension estimation. *Transactions of the American Matehmatical Society*, 353(11), 2001.

K. Chang and H. Tong. *Chaos: A statistical perspective.* Springer, 2001.

C. Cutler. *Dimension Estimation and Models - A Review of the theory and estimation of fractal dimension.* World Scientific, 1993.

W.D. Dechert. The correlation integral and the independece of stochastic processes. *Working paper*, 2003.

H. Deheling. Limit theorems for dependet U-statistics.

M. Denker and G. Keller. On U-statistics and V. Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gabiete*, 64:505–522, 1983.

C. Diks. Estimating invariants of noisy attractors. *Physical Review E*, 53(5):R4263–R4266, May 1996.

C. Diks. The correlation dimension in returns with stochastic volatility. *Quantitative Finance*, 4:45–54, 2004.

C. Diks. *In nonlinear time series analysis Methods and Application.* 1999.

G.A. Edgar. *Measure, Topology and Fractal Geometry.* Springer-Verlag, 1990.

M. Frank, G. Keller, and R. Sporer. Practical implementation of error estimation for the correlation dimension. *Physical Review E*, 53(6):5831–5836, June 1996.

P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9:189–208, 1983.

P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schriber. On noise reduction methods fro chaotic data. *CHAOS*, 3(2):127, 1993.

A. Guerrero and L.A. Smith. Towards coherent estiamtion of correlation dimension. *Physics Letters A*, 318, 2003.

K.P. Harikrishnan, R. Misra, G. Ambika, and A.K. Kembhavi. A non-subjective approach to the gp algorithm for analysing noisy time series. *Phisica D*, 215:137–145, 2006.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

T. Hsing. On tail index estimation using dependent data. *The Annals of Statistics*, 19 (3):1547–1569, 1991.

G. Keller and M. Kunzle. A maximum likelihood approach to dimension estimation. *preprint*, 1992.

G. Keller and R. Sporer. Remarks on the linear regression approach to dimension estimation.

E. Kocenda. An alternative to the bds test: integration across the correlation integral. *Econometric Reviews*, 20(3), 2001.

E. Kocenda and L. Briatka. Optimal range for the iid test based on integration across the correlation integral. *Econometric Reviews*, 24(3):265–296, 2005.

E.J. Kostelich and T. Schriber. Noise reduction in chaotic time-series data: A survey of commom methods. *Physical Review E*, 48, 1993.

G. Nolte, A. Ziehe, and K.R. Muller. Noise estimates of correlation dimension an k_2 entropy. *Physical Review E*, 64(016112):1–10, June 2001.

D.V. Pisarenko and V.F. Pisarenko. Statistical estimation of the correlation dimension. *Physics Letters A*, 197:31–39, 1995.

T. Schriber. Determination of the noise level of a chaotic time series. *Physical Review E*, 48(1):R13, 1993.

J. Serinko. A consistent approach to least squares estimation of correlation dimension in weak bernoulli dynamical system. *The Annals of Applied Probability*, 4(4), 1994.

J.C. Shouten, F. Takens, and C. van den Bleek. Estimation of the dimension of the noisy attractor. *Physical Review E*, 50(3):1851, September 1994.

R.L. Smith. Estimating dimension in noisy chaotic time series. *J.R. Statist. Soc. B*, 52: 329, 1992.

J. Theiler. Statistical precision of dimension estiamtors. *Physical Review A*, 41(6): 3038–3051, March 1990.

J. Theiler and T. Lookman. Statistical error in chord estimator of the correlation dimension: the rule of five. *Int. J. Bif. Chaos*, 3:765–771, 1993.

K. Yoshihara. Limiting behavior of U-statistics for stationary, absoletely regular process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gabiete*, 35:237–252, 1976.

D. Yu, M. Smith, R.G. Harrison, and C. Diks. Efficient implementation of the gaussian kernel algorithm in estimating invarinats and noise level from noisy time series data. *Physical Review E*, 61(4):3750–3756, April 2000.