The Correlation Integral and the Independence of Stochastic Processes

by

W. Davis Dechert[1,2]

Revised: July 2003[3]

[1]Honorary Fellow, Department of Economics, University of Wisconsin

[2]I thank Professor W. A. Brock for many helpful conversations on this topic. I also wish to thank Professors S. Durlauffm C. Granger, W. Heller and H. White for their comments as well as the members of the Econometric Workshops at the University of California at San Diego, the University of Iowa and the University of Maryland.
E–mail: wdechert@ssc.wisc.edu

[3]This revision was prepared while the author was a Visiting Research Fellow at the University of Wisconsin. I thank both the Department of Economics and the Social Systems Research Institute for their support.

A B S T R A C T

A generalization of the correlation integral of Grassberger and Procaccia is used to develop a statistic that has the property that it is asymptotically zero if and only if the underlying Gaussian process is independent. The same implication also holds for certain related processes. It is shown that the stastistic is asymptotically normal for weakly dependent stationary processes. An example is given of a zero autocovariance process. It also applies to ARCH (autoregressive conditionally heteroskedastic) random variables.

1 Introduction

In this paper I present a result that relates a property of a statistic used in the detection of chaotic dynamics to the property of independence of stochastic processes. The correlation integral of Grassberger and Procaccia (1983a, 1983b) is a measure of the frequency with which patterns in the data are repeated. It is based on the fact that chaotic dynamics exhibit regularity at highly irregular frequencies. This comes from two properties of chaotic dynamics: there are a large number of unstable periodic points in the phase space, and the chaotic trajectories are dense in the phase space. Thus, a chaotic trajectory will pass arbitrarily close to a given period orbit infinitely often and trace out its pattern before wandering off to some other part of the phase space. It should come as no surprise that when the correlation integral picks up a large number of repeated patterns in a data set that we would reject the hypothesis that the data were generated by an independent and identically distributed (IID) stochastic process. This is in part the result derived by Brock, Dechert, and Scheinkman (1987) and Brock, Dechert, Scheinkman, and LeBaron (1996).

What happens when the correlation integral only picks up repeated patterns to the extent that they would be expected to appear in independent data? In this paper I show that for certain stationary processes (which include the Gaussian) a statistic based on a generalization of the correlation integral is asymptotically zero if and *only if* the underlying process is stochastically independent. Furthermore, the same holds for certain related processes, among them the zero autocovariance process. It also holds for ARCH (autoregressive conditionally heteroskedastic) random variables as well.

The correlation integral based on m-histories of a data set, $\{x_t\}$ is given by

$$I_{m,n}(\epsilon) = \binom{n}{2}^{-1} \sum\sum_{1 \le s < t \le n} \prod_{k=0}^{m-1} \chi_\epsilon(|x_{s+k} - x_{t+k}|) \tag{1.1}$$

where

$$\chi_\epsilon(a) = \begin{cases} 1 & \text{if } a < \epsilon \\ 0 & \text{if } a \ge \epsilon \end{cases}$$

The statistic that was studied in Brock, Dechert, and Scheinkman (1987) is

$$T_{m,n} = I_{m,n}(\epsilon) - I_{1,n}(\epsilon)^m \tag{1.2}$$

In Brock and Dechert (1988) it was shown that if a stochastic data sequence $\{x_t\}$ is IID then

$$\lim_{n\to\infty} T_{m,n} = 0 \qquad \text{wp } 1 \tag{1.3}$$

and in Brock, Dechert, and Scheinkman (1987) it was shown that under a technical condition (which is discussed in section 3) on the distribution of the random variables the statistic $\sqrt{n}T_{m,n}$ is asymptotically normal. The technique that was used by them can also be adapted to other functions of the correlation integral such as an estimate for the dimension of the data set,

$$\hat{d}_{m,n}(\epsilon) = \frac{\ln I_{m,n}(\epsilon)}{\ln \epsilon} \tag{1.4}$$

for small values of ϵ.

Monte Carlo studies in Brock, Dechert, Scheinkman, and LeBaron (1996) and Brock, Hsieh, and LeBaron (1991) demonstrate that the statistic $T_{m,n}$ is useful for testing the null hypothesis of independently generated data against a wide variety of alternatives, and in particular against many distributions for which the data are stochastically uncorrelated, but not independent. In Brock, Dechert, and Scheinkman (1987) it is argued that the limit in equation (1.3) almost implies that the data are independently distributed.

The results in this paper follow from the fact that limits of the type in equation (1.3) impose restrictions on the distribution function of the data. (The restriction is actually derived in terms of the characteristic function of the generating process.) These restrictions can be used to determine whether or not the underlying process is stochastically independent.

2 Independence and the Correlation Integral

In some of the literature on nonlinear dynamics, the correlation integral of Grassberger and Procaccia (1983a, 1983b) in (1.1) has been used to test data for the presence of chaos, and in particular it has been used to estimate the dimension of data sets by using equation (1.4). The correlation integral has also been used in Brock, Dechert, and Scheinkman (1987) and Brock, Dechert, Scheinkman, and LeBaron (1996) to test time series for the presence of nonlinear stochastic dependence. The tests which were developed there are particularly effective for testing the null hypothesis of IID data against a variety of alternative distributions, including distributions for which the data are serially uncorrelated.

Based on the idea of the correlation integral, define the following: for $\epsilon, \epsilon_1, \epsilon_2 > 0$,

$$\begin{aligned} C_{X_t}(\epsilon) &= \iint_{x-\epsilon}^{x+\epsilon} dF_{X_t}(y) dF_{X_t}(x) \\ C_{X_s,X_t}(\epsilon_1,\epsilon_2) &= \iiint_{x_1-\epsilon_1}^{x_1+\epsilon_1} \int_{x_2-\epsilon_2}^{x_2+\epsilon_2} dF_{X_s,X_t}(y_1,y_2) dF_{X_s,X_t}(x_1,x_2) \end{aligned} \tag{2.1}$$

and extend these definitions to make them odd functions:

$$\begin{array}{rcll} C_{X_t}(\epsilon) &=& -C_{X_t}(-\epsilon) & \epsilon < 0 \\ C_{X_s,X_t}(\epsilon_1,\epsilon_2) &=& -C_{X_s,X_t}(-\epsilon_1,\epsilon_2) & \epsilon_1 < 0,\ \epsilon_2 > 0 \\ C_{X_s,X_t}(\epsilon_1,\epsilon_2) &=& -C_{X_s,X_t}(\epsilon_1,-\epsilon_2) & \epsilon_1 > 0,\ \epsilon_2 < 0 \\ C_{X_s,X_t}(\epsilon_1,\epsilon_2) &=& C_{X_s,X_t}(-\epsilon_1,-\epsilon_2) & \epsilon_1 < 0,\ \epsilon_2 < 0 \end{array}$$

These functions and the characteristic functions of the sequence $\{X_t\}$ satisfy the following:

Lemma 1 $\forall(\epsilon_1,\epsilon_2)$ $C_{X_s,X_t}(\epsilon_1,\epsilon_2) = C_{X_s}(\epsilon_1)C_{X_t}(\epsilon_2)$ *if and only if*

$$|\hat{F}_{X_s,X_t}(u_1,u_2)|^2 + |\hat{F}_{X_s,X_t}(u_1,-u_2)|^2 = 2|\hat{F}_{X_s}(u_1)|^2|\hat{F}_{X_s}(u_2)|^2 \tag{2.2}$$

where $\hat{F}_{X_s,X_t}(u_1,u_2)$ *is the characteristic function of* $F_{X_s,X_t}(x_1,x_2)$*:*

$$\hat{F}_{X_s,X_t}(u_1,u_2) = \iint e^{iu_1x_1} e^{iu_2x_2} dF_{X_s,X_t}(x_1,x_2)$$

and similarly for $\hat{F}_{X_t}(u)$:

$$\hat{F}_{X_t}(u) = \int e^{iux} dF_{X_t}(x).$$

Proof: All proofs are in the Appendix.

Notice that if X_s and X_t are independent then the condition in (2.2) is automatically satisfied, while the converse need not hold. However, for some distributions—including the Gaussian—condition (2.2) does imply that the random variables X_s and X_t are independent.

Theorem 2 *If X_s and X_t are jointly Gaussian and* $\forall(\epsilon_1, \epsilon_2)$

$$C_{X_s,X_t}(\epsilon_1, \epsilon_2) = C_{X_s}(\epsilon_1) C_{X_t}(\epsilon_2), \tag{2.3}$$

then X_s and X_t are independent. (Throughout, it is implicitly assumed that the random variables $\{X_t\}$ are not degenerate.)

Corollary 3 *If $\{X_t\}$ is a family of Gaussian random variables and* $(\forall s \neq t)(\forall \epsilon_1, \epsilon_2)$

$$C_{X_s,X_t}(\epsilon_1, \epsilon_2) = C_{X_s}(\epsilon_1) C_{X_t}(\epsilon_2)$$

then $\{X_t\}$ is a sequence of independent random variables.

Another distribution that is also independent if condition (2.2) holds is the Gaussian conditional heteroskedastic distribution which is the basis of the ARCH time series model developed by Engle (1982):

Corollary 4 *If the joint distribution of X_1, X_2 is*

$$\frac{\exp\left\{-\frac{x_1^2}{2} - \frac{x_2^2}{2(\alpha+\beta x_1^2)}\right\}}{(2\pi)(\alpha + \beta x_1^2)^{\frac{1}{2}}}$$

where $\alpha > 0$ and $\alpha + \beta = 1$ then condition (2.2) holds if and only if $\beta = 0$.

In time series analysis, it is often assumed that the data $\{X_t\}$ forms a stationary sequence. In this case define

$$\begin{aligned} C(\epsilon) &= C_{X_t}(\epsilon) \\ C_m(\epsilon_1, \epsilon_2) &= C_{X_t,X_{t+m}}(\epsilon_1, \epsilon_2) \end{aligned} \tag{2.4}$$

which are independent of t by stationarity. Then,

Corollary 5 *If $\{X_t\}$ is a sequence of stationary Gaussian random variables and* $(\forall \epsilon_1, \epsilon_2)$

$$C_m(\epsilon_1, \epsilon_2) = C(\epsilon_1) C(\epsilon_2) \tag{2.5}$$

then X_t and X_{t+m} are independent for all t. Furthermore, if (2.5) holds for all $m \geq 1$, then the sequence $\{X_t\}$ is IID.

Certain types of time series can derived from a stationary Gaussian process. For example consider the zero autocovariance process:

$$x_t = \epsilon_t + \alpha\epsilon_{t-1}\epsilon_{t-2}$$

where $\{\epsilon_t\}$ is an iid Gaussian process with zero mean and unit variance. Then the $\{x_t\}$ process has zero mean and variance $1+\alpha^2$. It also has the property that $\mathsf{E}\left[x_s x_t\right] = 0$ for $s \neq t$ and yet they are not independent if $|t-s| \leq 2$, unless $\alpha = 0$. This process also has the property that condition (2.2) is satisfied if and only if $\alpha = 0$, i.e., the sequence is IID.

Corollary 6 *If the zero autocovariance process satisfies*

$$C_m(\epsilon_1, \epsilon_2) = C(\epsilon_1)C(\epsilon_2)$$

for $m = 1, 2$ and for all ϵ_1, ϵ_2 then $\alpha = 0$.

Theorem 2 and its corollaries provide a characterization of independence for Gaussian random variables in terms of the correlation integral like functions in (2.1). The methods and results of this paper can be readily applied to a generalized version of the correlation integral:

$$C_{X_{t_1},\ldots,X_{t_m}}(\epsilon_1,\ldots,\epsilon_m) =$$
$$\int\cdots\int \chi_{\epsilon_1}(|x_1-y_1|)\ldots\chi_{\epsilon_m}(|x_m-y_m|)dF_{X_{t_1},\ldots,X_{t_m}}(x_1,\ldots,x_m)dF_{X_{t_1},\ldots,X_{t_m}}(y_1,\ldots,y_m).$$

Then

$$C_{X_{t_1},\ldots,X_{t_m}}(\epsilon_1,\ldots,\epsilon_m) = C_{X_{t_1}}(\epsilon_1)\ldots C_{X_{t_m}}(\epsilon_m)$$

for all $(\epsilon_1,\ldots,\epsilon_m)$ if and only if

$$\begin{aligned}|\hat{F}_{X_{t_1},X_{t_2},\ldots,X_{t_m}}(u_1,u_2,\ldots,u_m)|^2 + |\hat{F}_{X_{t_1},X_{t_2},\ldots,X_{t_m}}(-u_1,u_2,\ldots,u_m)|^2 + \ldots \\ = 2^m|\hat{F}_{X_{t_1}}(u_1)|^2|\hat{F}_{X_{t_2}}(u_2)|^2\cdots|\hat{F}_{X_{t_m}}(u_m)|^2\end{aligned} \tag{2.6}$$

where the left hand side of equation (2.6) is summed over all of the 2^m possible sign patterns of $(\pm u_1, \pm u_2, \ldots, \pm u_m)$.

Condition (2.2), or more generally condition (2.6), is the key to the link between the factoring of the generalized correlation integral and the independence of the underlying process. For those processes for which this condition implies that the characteristic function factors, we have the result that the correlation integral factors if and only if the process is independent.

3 Testing Time Series for Independence

Let $\{X_t\}$ be a stationary time series, and define

$$C_{m,n}(\epsilon_1,\epsilon_2) = \binom{n}{2}^{-1} \sum\sum_{1\leq s<t\leq n} \chi_{\epsilon_1}(|x_s-x_t|)\chi_{\epsilon_2}(|x_{s+m}-x_{t+m}|) \tag{3.1}$$

for $m \geq 2$, and

$$C_n(\epsilon) = \binom{n}{2}^{-1} \sum\sum_{1 \leq s < t \leq n} \chi_\epsilon(|x_s - x_t|) \tag{3.2}$$

The statistic in (3.1) is a generalized U-statistic (see Serfling (1980), chapter 5). When the random variables, $\{X_t\}$, are independent then

$$\mathsf{E}[\chi_\epsilon(|X_s - X_t|)] = C(\epsilon) \tag{3.3}$$

and

$$\begin{aligned} \mathsf{E}[C_n(\epsilon)] &= C(\epsilon) &(3.4)\\ \lim_{n\to\infty} \mathsf{E}[C_{m,n}(\epsilon_1,\epsilon_2)] &= C(\epsilon_1)C(\epsilon_2) \quad \text{wp 1.} &(3.5)\end{aligned}$$

Thus the statistic

$$S_{m,n}(\epsilon_1,\epsilon_2) = C_{m,n}(\epsilon_1,\epsilon_2) - C_n(\epsilon_1)C_n(\epsilon_2) \tag{3.6}$$

has $\lim_{n\to\infty} S_{m,n} = 0$ almost surely.

By the delta method (see Pollard (1984), Appendix A), the limit distribution of $\sqrt{n}S_{m,n}(\epsilon_1,\epsilon_2)$ is the same as the limit distribution of

$$\sqrt{n}\left\{ [C_{m,n}(\epsilon_1,\epsilon_2) - C(\epsilon_1)C(\epsilon_2)] - C(\epsilon_2)\left[C_n(\epsilon_1) - C(\epsilon_1)\right] - C(\epsilon_1)\left[C_n(\epsilon_2) - C(\epsilon_2)\right]\right\}. \tag{3.7}$$

The term in braces in (3.7) is a generalized U-statistic with kernel

$$\begin{aligned} h(\boldsymbol{x},\boldsymbol{y}) &= h(x_1,x_2,y_1,y_2)\\ &= \chi_{\epsilon_1}(|x_1-y_1|)\chi_{\epsilon_2}(|x_2-y_2|)\\ &\quad -C(\epsilon_2)\chi_{\epsilon_1}(|x_1-y_1|) - C(\epsilon_1)\chi_{\epsilon_1}(|x_2-y_2|), \end{aligned} \tag{3.8}$$

and following the notation of Denker and Keller (1983) and Serfling (1980), define

$$\begin{aligned} h_1(\boldsymbol{x}) &= h_1(x_1,x_2) = \iint h(x_1,x_2,y_1,y_2)dF(y_1)dF(y_2)\\ &= [F(x_1+\epsilon_1) - F(x_1-\epsilon_1)][F(x_2+\epsilon_2) - F(x_2-\epsilon_2)]\\ &\quad -C(\epsilon_2)[F(x_1+\epsilon_1) - F(x_1-\epsilon_1)] - C(\epsilon_1)[F(x_2+\epsilon_2) - F(x_2-\epsilon_2)]\\ &\quad +C(\epsilon_1)C(\epsilon_2). \end{aligned}$$

The asymptotic variance of the statistic in (3.7) is given by

$$\sigma^2 = 4\left\{\mathsf{E}[h_1(X_1,X_2)^2] + 2\mathsf{E}[h_1(X_1,X_2)h_1(X_2,X_3)]\right\}.$$

If we define $K(\epsilon) = \mathsf{E}[(F(X_1+\epsilon) - F(X_1-\epsilon))^2]$ then

$$\sigma^2 = 4[K(\epsilon_1) - C(\epsilon_1)^2][K(\epsilon_2) - C(\epsilon_2)^2].$$

These details are used in the following theorem on the asymptotic behavior of the statistic $S_{m,n}$.

Theorem 7 *If $K(\epsilon_i) - C(\epsilon_i)^2 > 0$ for $i = 1, 2$ and if the random variables $\{X_t\}$ are independent, then*

$$\sqrt{n}\frac{S_{m,n}(\epsilon_1, \epsilon_2)}{\sigma(\epsilon_1, \epsilon_2)} \xrightarrow{D} N(0, 1). \tag{3.9}$$

The asymptotic variance can be consistently estimated by:

$$\sigma_n^2(\epsilon_1, \epsilon_2) = 4[\hat{K}_n(\epsilon_1) - \hat{C}_n(\epsilon_1)^2][\hat{K}_n(\epsilon_2) - \hat{C}_n(\epsilon_2)^2]$$

where

$$\hat{K}_n(\epsilon) = \frac{1}{n^3}\sum_{r=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\chi_\epsilon(|x_r - x_s|)\chi_\epsilon(|x_s - x_t|) \tag{3.10}$$

$$= \frac{1}{n^3}\sum_{r=1}^{n}\left(\sum_{s=1}^{n}\chi_\epsilon(|x_r - x_s|)\right)^2 \tag{3.11}$$

and

$$\hat{C}_n(\epsilon) = \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\chi_\epsilon(|x_s - x_t|). \tag{3.12}$$

The formula in equation (3.11) is useful for speeding up computer calculations.

The technical condition, $K - C^2 > 0$ implies both the $\sqrt{n}$ scaling in (3.9) as well as the asymptotic normality result. When $K - C^2 = 0$, neither of these hold. Hölder's inequality shows when this latter case applies:

$$\begin{aligned} C(\epsilon) &= \int [F(x+\epsilon) - F(x-\epsilon)]dF(x) \\ &\leq \left(\int [F(x+\epsilon) - F(x-\epsilon)]^2 dF(x)\right)^{\frac{1}{2}} \left(\int 1^2 dF(x)\right)^{\frac{1}{2}} = K(\epsilon)^{\frac{1}{2}} \end{aligned}$$

with equality if and only if there is a constant γ such that for all x

$$F(x+\epsilon) - F(x-\epsilon) = \gamma. \tag{3.13}$$

The asymptotic convergence theorem and the calculation of the variance follow from the methods of either Denker and Keller (1983) for absolutely regular processes, or Sen (1963) for m-dependent processes. The kernel for $C_{m,n}$ is a function of m-histories of the data, $x_t^m = (x_t, x_{t+1}, \ldots, x_{t+m-1})$, and the random vectors X_s^m and X_t^m are not independent unless $|t - s| > m$. The sequence $\{X_t^m\}$ is m-dependent, which is a special case of an absolutely regular process.

In a Monte Carlo study Theiler (1988) showed that $S_{2,n}(\epsilon, \epsilon)/\sigma_n(\epsilon, \epsilon)$ does not scale like $1/\sqrt{n}$ but rather like $1/n$ when the data are IID uniform on the interval $[0, 1]$ and a wrap-around metric is used:

$d(x,y) = \min\{|x-y|, 1-|x-y|\}$. This distribution is equivalent to a uniform distribution on the unit circle. This is the only degenerate case. For this case it turns out that $F(x+\epsilon) - F(x-\epsilon) = 2\epsilon$ for all $x \in [0,1]$, and so it satisfies the Hölder condition (3.13). For a discussion of the statistical properties of $S_{m,n}$ when $K - C^2 = 0$ see Serfling (1980, Section 5.3.4.). This case has an impact on the statistic for small data sets when the data is uniformly distributed on an interval since $F(x+\epsilon) - F(x-\epsilon)$ is constant over a large part of the interval for small ϵ. See Brock, Hsieh, and LeBaron (1991) for Monte Carlo studies.

4 Rates of Convergence and Choice of ϵ

The correlation dimension of a data set $\{x_t\}$ at embedding dimension m is given by

$$I_m = \lim_{n\to\infty} I_{m,n}(\epsilon) \tag{4.1}$$

$$d_m = \limsup_{\epsilon\to 0} \frac{\ln I_m(\epsilon)}{\ln \epsilon}. \tag{4.2}$$

When this is estimated on a finite data set by (1.4), the choice of ϵ plays an important role. Typical log–log plots of $\hat{d}_{m,n}(\epsilon)$ 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 d_m. This experimental procedure requires an additional step in the estimation of (4.2), namely the estimation of the slope of the graph of (1.4). This step introduces a bias in the estimate of the dimension of the data set which is discussed in Ramsey and Yuan (1987).

In this section, I propose a systematic way of choosing ϵ which is based on the statistical properties of U-statistics. It is a method for choosing a sequence $\{\epsilon_n\}$ such that $\epsilon_n \to 0$ and for which the statistical results of Section 3 still hold. The distribution function for the U-statistic $I_{1,n}(\epsilon)$ in (1.1) and $C_{m,n}(\epsilon)$ in (3.1) is:

$$F_n(t;\epsilon) = \mathcal{P}\left(\sqrt{n}\frac{I_{1,n}(\epsilon) - C(\epsilon)}{2\sqrt{K(\epsilon) - C(\epsilon)^2}} < t\right)$$

and the following result of Callaert and Janssen (1978) provides the rate of convergence of this distribution to the Normal:

Theorem 8 *If $K(\epsilon) - C(\epsilon)^2 > 0$ then $\exists \alpha < \infty$*

$$||F_n(\cdot;\epsilon) - \Phi||_\infty = \sup_t |F_n(t;\epsilon) - \Phi(t)| \le \frac{\alpha C(\epsilon)}{\sqrt{n}[K(\epsilon) - C(\epsilon)^2]^{\frac{3}{2}}}$$

where Φ is the standard Normal distribution function.

A consequence of this theorem is that if ϵ_n converges slowly enough to 0 so that $n[K(\epsilon_n) - C(\epsilon_n)^2]^3/C(\epsilon_n)^2$ diverges, then $\lim_{n\to\infty} ||F_n(\cdot;\epsilon_n) - \Phi||_\infty = 0$.

Two examples will provide some insight into the rate at which ϵ can be taken to 0.

(i) The Uniform distribution, $F(x) = x$, on the interval $[0, 1]$:

$$\begin{aligned} C(\epsilon) &= \int_0^1 [F(x+\epsilon) - F(x-\epsilon)]dx = 2\epsilon - \epsilon^2 \\ K(\epsilon) &= \int_0^1 [F(x+\epsilon) - F(x-\epsilon)]^2 dx = 4\epsilon^2 - 3\epsilon^3 \end{aligned}$$

and so $K(\epsilon) - C(\epsilon)^2 = \epsilon^3(1-\epsilon)$. Thus $[K(\epsilon) - C(\epsilon)^2]^3/C(\epsilon)^2 = O(\epsilon^7)$, and the sequence $\{\epsilon_n\}$ has to be chosen so that $n\epsilon_n^7 \to \infty$.

(ii) The Exponential distribution, $F(x) = 1 - e^{-\lambda x}$, on the interval $(0, \infty)$:

$$\begin{aligned} C(\epsilon) &= \frac{1}{2}[e^{\lambda\epsilon} - e^{-\lambda\epsilon}] \\ K(\epsilon) &= \frac{1}{3}[e^{\lambda\epsilon} - e^{-\lambda\epsilon}]^2 \end{aligned}$$

and $K(\epsilon) - C(\epsilon)^2 = (e^{\lambda\epsilon} - e^{-\lambda\epsilon})^2/12$. Therefore $[K(\epsilon) - C(\epsilon)^2]^3/C(\epsilon)^2 = (e^{\lambda\epsilon} - e^{-\lambda\epsilon})^4/432 = O(\epsilon^4)$ and the sequence $\{\epsilon_n\}$ must satisfy $n\epsilon_n^4 \to \infty$.

The next result shows that $I_{1,n}(\epsilon) \to C(\epsilon)$ wp 1 uniformly in ϵ. This in turn implies that $I_{1,n}(\epsilon_n) - C(\epsilon_n) \to 0$ for any convergent sequence $\{\epsilon_n\}$.

Theorem 9 *Let $\{X_t\}$ be IID. Then*

$$\mathcal{P}\left(\lim_{n\to\infty} \sup_\epsilon |I_{1,n}(\epsilon) - C(\epsilon)| = 0\right) = 1.$$

Notice that although Theorem 9 is stated for $\{X_t\}$ IID, it holds under the same conditions as the main theorem of Denker and Keller (1983, p. 506). Furthermore, it also holds when the data $\{x_t\}$ are generated by a mapping f,

$$x_{t+1} = f(x_t)$$

where f has an invariant measure, ρ, on an indecomposable attractor. The only step in the theorem that uses these facts is that $\mathcal{P}(A_{ij}) = 1$, which follows from $I_{1,n}(\epsilon) \to C(\epsilon)$ wp 1. This conclusion remains valid if the probability measure $\mathcal{P}$ is replaced by the ergodic measure ρ.

5 Deterministic Data

Suppose that the data are generated by a deterministic system,

$$x_{t+1} = f(x_t) \tag{5.1}$$

where I will focus on the case that $f : R \to R$. A subset $\Lambda \subset R$ is called an indecomposable attractor if it is invariant, $f(\Lambda) = \Lambda$, if no proper subsets of Λ are invariant, and if there is an open set U containing Λ such that for all $x_0 \in U$, the orbit $\{x_t\}$ converges to Λ. I will also assume that there is an ergodic measure, ρ, for f on Λ.

A simple yet illustrative example is the tent map, $T(x) = 1 - |1 - 2x|$. The interval $(0, 1)$ is an attractor, and the ergodic measure coincides with Lebesgue measure. Furthermore, for almost all $x_0 \in (0, 1)$ the orbit starting from x_0 has the appearance of white noise, in that the correlation coefficients are all zero.

For deterministic data (such as from the tent map above) the statistics of section 3 behave differently than they do for stochastic data:

Lemma 5.1 *Assume that f satisfies a Lipschitz condition:* $(\exists k)(\exists \alpha)(\forall x, y \in \Lambda) \quad |f(x) - f(y)| \leq k|x - y|^\alpha$. *If $\epsilon_2 \geq k^{1+\alpha+\cdots+\alpha^{m-1}} \epsilon_1^{\alpha^m}$ then*

$$C_m(\epsilon_1, \epsilon_2) = C(\epsilon_1)$$

and

$$C_{m,n}(\epsilon_1, \epsilon_2) = C_n(\epsilon_1).$$

Notice that when the conditions of the lemma hold then

$$S_{m,n}(\epsilon, k^{1+\alpha+\cdots+\alpha^{m-1}} \epsilon^{\alpha^m}) = C_n(\epsilon)[1 - C_n(k^{1+\alpha+\cdots+\alpha^{m-1}} \epsilon^{\alpha^m})],$$

and for values of ϵ, k, α such that $S_{m,n} \neq 0$, then

$$\lim_{n\to\infty} S_{m,n}(\epsilon, k^{1+\alpha+\cdots+\alpha^{m-1}} \epsilon^{\alpha^m}) = C(\epsilon)[1 - C(k^{1+\alpha+\cdots+\alpha^{m-1}} \epsilon^{\alpha^m})] \tag{5.2}$$

and

$$\lim_{n\to\infty} \sqrt{n} S_{m,n}(\epsilon, k^{1+\alpha+\cdots+\alpha^{m-1}} \epsilon^{\alpha^m}) = \infty. \tag{5.3}$$

Recall that for stochastic data, the limit in (5.2) is zero and the limit in (5.3) is a Gaussian random variable. This yields a procedure that can separate stationary random systems with independent observations from deterministic systems which are ergodic. For other aspects of this issue, see Brock (1986) and Brock and Dechert (1988).

When there is measurement noise in the system,

$$\begin{aligned} y_{t+1} &= f(y_t) \\ x_t &= y_t + \sigma w_t \end{aligned} \tag{5.4}$$

(where $\{w_t\}$ are IID with $\mathsf{E}w_t = 0$ and $\mathsf{E}w_t^2 = 1$) then for large σ one would expect that the sequence $\{x_t\}$ would behave like a random sequence and that $S_{m,n}(\epsilon, k^{1+\alpha+\cdots+\alpha^{m-1}}\epsilon^{\alpha^m})$ would converge to zero. For small σ one would expect that the sequence would behave like deterministic data and that $S_{m,n}(\epsilon, k^{1+\alpha+\cdots+\alpha^{m-1}}\epsilon^{\alpha^m})$ would converge to a positive constant. The following theorem provides the continuity property of $C_m(\epsilon_1, \epsilon_2)$:

Theorem 5.2 *Let the data be generated by model (5.4). Then for almost all initial conditions, y_0, and for $\epsilon_2 \geq k^{1+\alpha+\cdots+\alpha^{m-1}}\epsilon_1^{\alpha^m}$*

$$\lim_{\sigma\to\infty}[C_m(\epsilon_1, \epsilon_2) - C(\epsilon_1)C(\epsilon_2)] = 0$$

and

$$\lim_{\sigma\to 0}[C_m(\epsilon_1, \epsilon_2) - C(\epsilon_1)] = 0$$

ρ–almost everywhere.

If we define

$$\sigma_f = \int x^2 d\rho(x) - \left(\int x d\rho(x)\right)^2$$

then the signal to noise ratio is σ_f/σ, and theorem 5.2 shows that as the signal to noise ratio increases, the data generated by the system behaves more like deterministic data than random data.

In Figure 1 there are plots for a tent map where the observation errors are IID Gaussian random variables. Signal to noise ratios of 1, 10, and 100 were used. For a signal to noise ratio of 100, the value of the statistic agrees quite closely with the theoretical value of $C(\epsilon)[1 - C(2\epsilon)] = 0.12$. The actual values are in Table 1. Notice in Figure 1 that the slope of the upper graph ($\sigma_f/\sigma = 10$) is 1/2 for all values of n, showing that for tent map data with a small amount of noise the statistic converges rapidly to a constant times $\sqrt{n}$. Even for a low signal to noise ratio ($\sigma_f/\sigma = 1$) the statistic picks up the presence of non-linear structure for $n \geq 4000$.

n	$\sigma_f/\sigma = 1$	$\sigma_f/\sigma = 10$	$\sigma_f/\sigma = 100$
109	1.911×10^{-3}	9.518×10^{-2}	1.144×10^{-1}
148	2.661×10^{-3}	9.451×10^{-2}	1.157×10^{-1}
200	1.648×10^{-3}	9.348×10^{-2}	1.166×10^{-1}
270	1.634×10^{-3}	9.441×10^{-2}	1.184×10^{-1}
365	1.319×10^{-3}	9.509×10^{-2}	1.193×10^{-1}
492	1.345×10^{-3}	9.494×10^{-2}	1.202×10^{-1}
665	2.750×10^{-3}	9.460×10^{-2}	1.198×10^{-1}
897	2.003×10^{-3}	9.405×10^{-2}	1.195×10^{-1}
1211	1.814×10^{-3}	9.298×10^{-2}	1.186×10^{-1}
1635	1.837×10^{-3}	9.245×10^{-2}	1.185×10^{-1}
2208	1.482×10^{-3}	9.230×10^{-2}	1.186×10^{-1}
2980	1.285×10^{-3}	9.172×10^{-2}	1.185×10^{-1}
4023	1.217×10^{-3}	9.216×10^{-2}	1.189×10^{-1}
5431	1.294×10^{-3}	9.227×10^{-2}	1.190×10^{-1}
7331	1.379×10^{-3}	9.231×10^{-2}	1.191×10^{-1}

Table 1: Values of $S_{1,n}(\epsilon, 2\epsilon)$ for the tent map

6 Conclusion

In this paper I have presented a variant of the correlation integral on 2-histories which can be used to test whether or not a stationary times series is IID. One aspect of interest of this test is that for those processes for which condition (2.2) — or more generally, condition (2.6) — holds is that independence is equivalent to the condition that

$$\mathsf{P}\left\{|X_s - X_t| < \epsilon_1 \wedge |X_{s+m} - X_{t+m}| < \epsilon_2\right\}$$

factors into the product

$$\mathsf{P}\left\{|X_s - X_t| < \epsilon_1\right\} \mathsf{P}\left\{|X_{s+m} - X_{t+m}| < \epsilon_2\right\}.$$

Thus, independence for these processes involves the behavior of the random variables along the diagonal. As it was shown in section 2, the Gaussian distribution (and some related distributions) are among those that satisfy the conditions of this paper.

In terms of the BDS statistic of Brock, Dechert, and Scheinkman (1987), condition (2.6) can be used to determine those distributions for which the test of IID has no power. Since the BDS test has been widely used in recent years,[4] this should provide some additional insight as to its applicability.

[4]For a review of recent results in the application of the BDS test, see Brock, Dechert, Scheinkman, and LeBaron (1996).

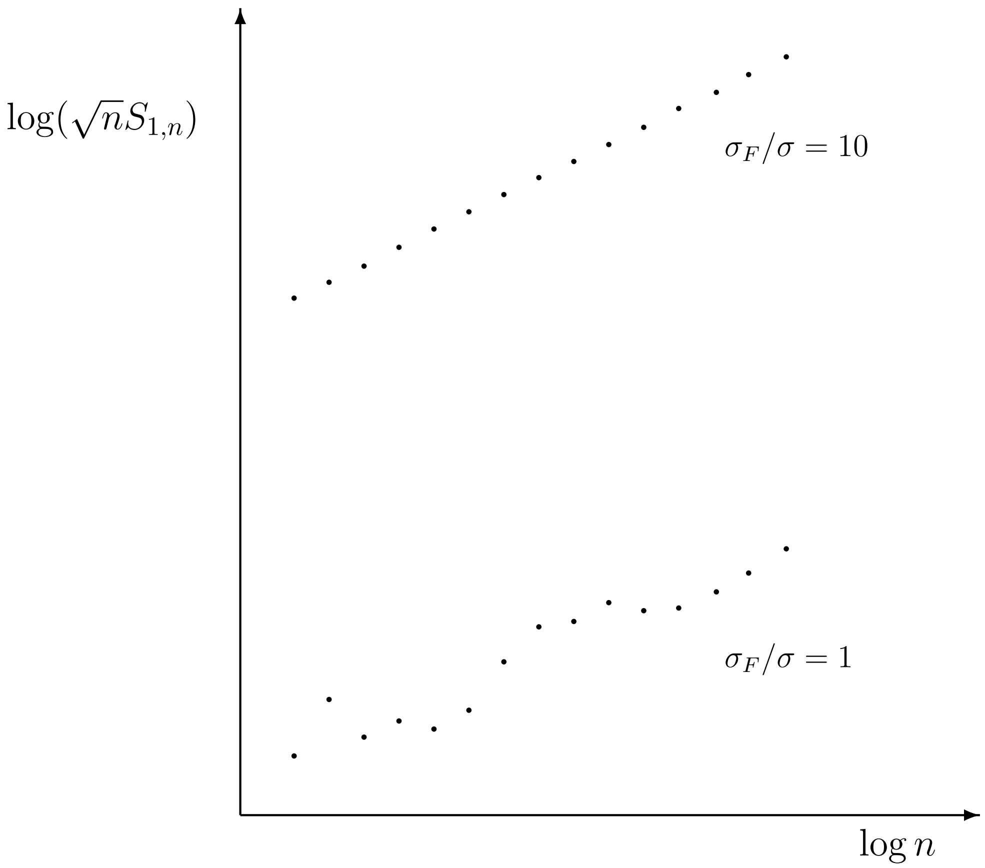

Figure 1

Appendix

Lemma 1: $\forall(\epsilon_1, \epsilon_2)$ $C_{X_s,X_t}(\epsilon_1, \epsilon_2) = C_{X_s}(\epsilon_1)C_{X_t}(\epsilon_2)$ *if and only if*

$$|\hat{F}_{X_s,X_t}(u_1, u_2)|^2 + |\hat{F}_{X_s,X_t}(u_1, -u_2)|^2 = 2|\hat{F}_{X_s}(u_1)|^2|\hat{F}_{X_s}(u_2)|^2 \tag{A.1}$$

where $\hat{F}_{X_s,X_t}(u_1, u_2)$ *is the characteristic function of* $F_{X_s,X_t}(x_1, x_2)$*:*

$$\hat{F}_{X_s,X_t}(u_1, u_2) = \iint e^{iu_1x_1}e^{iu_2x_2}dF_{X_s,X_t}(x_1, x_2)$$

and similarly for $\hat{F}_{X_t}(u)$.

Proof:

$$\begin{aligned}C_{X_s,X_t}(\epsilon_1,\epsilon_2) = \iint &[F_{X_s,X_t}(x_1+\epsilon_1,x_2+\epsilon_2) - F_{X_s,X_t}(x_1+\epsilon_1,x_2-\epsilon_2)\\ &-F_{X_s,X_t}(x_1-\epsilon_1,x_2+\epsilon_2) + F_{X_s,X_t}(x_1-\epsilon_1,x_2-\epsilon_2)\,]\,dF_{X_s,X_t}(x_1,x_2)\end{aligned}$$

which implies that

$$\begin{aligned}\hat{C}_{X_s,X_t}(u_1,u_2) &= \iint e^{iu_1x_1}e^{iu_2x_2}dC_{X_s,X_t}(\epsilon_1,\epsilon_2)\\ &= \iiiint e^{iu_1x_1}e^{iu_2x_2}dF_{X_s,X_t}(x_1+\epsilon_1,x_2+\epsilon_2)dF_{X_s,X_t}(x_1,x_2)\\ &\quad+\iiiint e^{iu_1x_1}e^{iu_2x_2}dF_{X_s,X_t}(x_1+\epsilon_1,x_2-\epsilon_2)dF_{X_s,X_t}(x_1,x_2)\\ &\quad+\iiiint e^{iu_1x_1}e^{iu_2x_2}dF_{X_s,X_t}(x_1-\epsilon_1,x_2+\epsilon_2)dF_{X_s,X_t}(x_1,x_2)\\ &\quad+\iiiint e^{iu_1x_1}e^{iu_2x_2}dF_{X_s,X_t}(x_1-\epsilon_1,x_2-\epsilon_2)dF_{X_s,X_t}(x_1,x_2).\end{aligned}$$

By a change of variables, $y_1 = x_1 - \epsilon_1$, $y_2 = x_2 - \epsilon_2$,

$$\begin{aligned}\iiiint &e^{iu_1x_1}e^{iu_2x_2}dF_{X_s,X_t}(x_1+\epsilon_1,x_2+\epsilon_2)dF_{X_s,X_t}(x_1,x_2)\\ &= \iiiint e^{iu_1(y_1-x_1)}e^{iu_2(y_2-x_2)}dF_{X_s,X_t}(y_1,y_2)dF_{X_s,X_t}(x_1,x_2)\\ &= \hat{F}_{X_s,X_t}(u_1,u_2)\hat{F}_{X_s,X_t}(-u_1,-u_2)\\ &= |\hat{F}_{X_s,X_t}(u_1,u_2)|^2\end{aligned}$$

since $\overline{\hat{F}_{X_s,X_t}(u_1,u_2)} = \hat{F}_{X_s,X_t}(-u_1,-u_2)$. By similar calculations,

$$\begin{aligned}\hat{C}_{X_s,X_t}(u_1,u_2) &= |\hat{F}_{X_s,X_t}(u_1,u_2)|^2 + |\hat{F}_{X_s,X_t}(u_1,-u_2)|^2\\ &\qquad+|\hat{F}_{X_s,X_t}(-u_1,u_2)|^2 + |\hat{F}_{X_s,X_t}(-u_1,-u_2)|^2\\ &= 2|\hat{F}_{X_s,X_t}(u_1,u_2)|^2 + 2|\hat{F}_{X_s,X_t}(u_1,-u_2)|^2\end{aligned}$$

since $\overline{\hat{F}_{X_s,X_t}(u_1,-u_2)} = \hat{F}_{X_s,X_t}(-u_1,u_2)$. Similarly,

$$C_{X_t}(\epsilon) = \int [F(x+\epsilon) - F(x-\epsilon)]dF_{X_t}(x)$$

which implies that

$$\begin{aligned}\hat{C}_{X_t}(u) &= \int e^{iu\epsilon}dC_{X_t}(\epsilon)\\ &= \iint e^{iu\epsilon}dF_{X_t}(x+\epsilon)dF_{X_t}(x) + \iint e^{iu\epsilon}dF_{X_t}(x-\epsilon)dF_{X_t}(x)\\ &= \iint e^{iu(y-x)}dF_{X_t}(y)dF_{X_t}(x) + \iint e^{iu(x-y)}dF_{X_t}(y)dF_{X_t}(x)\\ &= \hat{F}_{X_t}(u)\hat{F}_{X_t}(-u) + \hat{F}_{X_t}(-u)\hat{F}_{X_t}(u)\\ &= 2|\hat{F}_{X_t}(u)|^2.\end{aligned}$$

Therefore, $\hat{C}_{X_s,X_t}(u_1,u_2) = \hat{C}_{X_s}(u_1)\hat{C}_{X_t}(u_2)$ and hence (A.1) holds. By the Fourier inversion formula,

$$\hat{C}_{X_s,X_t}(u_1,u_2) = \hat{C}_{X_s}(u_1)\hat{C}_{X_t}(u_2)$$

if and only if

$$C_{X_s,X_t}(\epsilon_1,\epsilon_2) = C_{X_s}(\epsilon_1)\hat{C}_{X_t}(\epsilon_2)$$

and the conclusion of the lemma follows. □

Theorem 2: *If X_s and X_t are jointly Gaussian and $\forall(\epsilon_1,\epsilon_2)$*

$$C_{X_s,X_t}(\epsilon_1,\epsilon_2) = C_{X_s}(\epsilon_1)C_{X_t}(\epsilon_2), \tag{A.2}$$

then X_s and X_t are independent. (Throughout, it is implicitly assumed that the random variables $\{X_t\}$ are not degenerate.)

Proof: Let $\mathsf{E}X_s = \mu_s$, $\mathsf{E}X_t = \mu_t$ and let the variance-covariance matrix be

$$\Sigma = \begin{bmatrix} \sigma_s^2 & \rho\sigma_s\sigma_t \\ \rho\sigma_s\sigma_t & \sigma_t^2 \end{bmatrix}$$

Then by Lemma 1,

$$|e^{iu_s\mu_s}e^{iu_t\mu_t}e^{\frac{1}{2}[u_s\ u_t]\Sigma[u_s\ u_t]'}|^2 + |e^{iu_s\mu_s}e^{-iu_t\mu_t}e^{\frac{1}{2}[u_s\ -u_t]\Sigma[u_s\ -u_t]'}|^2 = |e^{iu_s\mu_s}e^{\frac{1}{2}u_s^2\mu_s^2}||e^{iu_t\mu_t}e^{\frac{1}{2}u_t^2\mu_t^2}|$$

which reduces to

$$e^{2u_su_t\rho\sigma_s\sigma_t} + e^{-2u_su_t\rho\sigma_s\sigma_t} = 2$$

which holds for all (u_s,u_t) if and only if $\rho = 0$. □

Corollary 3: *If $\{X_t\}$ is a family of Gaussian random variables and $\forall(s \neq t)\forall(\epsilon_1,\epsilon_2)$*

$$C_{X_s,X_t}(\epsilon_1,\epsilon_2) = C_{X_s}(\epsilon_1)C_{X_t}(\epsilon_2)$$

then $\{X_t\}$ is a sequence of independent random variables.

Proof: This follows immediately from the theorem, since if a family of normal random variables are pairwise independent, they are independent. □

Corollary 4: *If the joint distribution of X_1, X_2 is*

$$\frac{\exp\left\{-\frac{x_1^2}{2} - \frac{x_2^2}{2(\alpha+\beta x_1^2)}\right\}}{(2\pi)(\alpha+\beta x_1^2)^{\frac{1}{2}}}$$

where $\alpha > 0$ and $\alpha + \beta = 1$ then condition (2.2) holds if and only if $\beta = 0$.

Proof:

$$\iint \frac{\exp\left\{-\frac{x_1^2}{2} - \frac{x_2^2}{2(\alpha+\beta x_1^2)}\right\}}{(2\pi)(\alpha+\beta x_1^2)^{\frac{1}{2}}} e^{iu_1x_1} e^{iu_2x_2} dx_1 dx_2$$

$$= \int \frac{\exp\left\{-\frac{x_1^2}{2}\right\}}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}u_2^2(\alpha+\beta x_1^2)\right\} e^{iu_1x_1} dx_1$$

$$= \frac{\exp\left\{-\frac{1}{2}\alpha u_2^2\right\}}{\sqrt{1+\beta u_2^2}} \int \frac{\sqrt{(1+\beta u_2^2)} \exp\left\{-\frac{1}{2}(1+\beta u_2^2)x_1^2\right\}}{\sqrt{2\pi}} e^{iu_1x_1} dx_1$$

$$= \frac{\exp\left\{-\frac{1}{2}\alpha u_2^2\right\} \exp\left\{-\frac{1}{2}\frac{u_1^2}{1+\beta u_2^2}\right\}}{\sqrt{1+\beta u_2^2}}$$

and this equals

$$\frac{\exp\left\{-\frac{1}{2}\alpha u_2^2\right\}}{\sqrt{1+\beta u_2^2}} \exp\left\{-\frac{1}{2}u_1^2\right\}$$

if and only if $\beta = 0$. Thus condition (2.2) implies that the ARCH random variables are independent. □

Corollary 5: *If $\{X_t\}$ is a sequence of stationary Gaussian random variables and $\forall(\epsilon_1, \epsilon_2)$*

$$C_m(\epsilon_1, \epsilon_2) = C(\epsilon_1)C(\epsilon_2) \tag{A.3}$$

then X_t and X_{t+m} are independent for all t. Furthermore, if (A.3) holds for all $m \geq 1$, then the sequence $\{X_t\}$ is IID.

Proof: This follows from Theorem 2, Corollary 3 and the stationarity of the sequence $\{X_t\}$. □

Corollary 6: *If the zero autocovariance process satisfies*

$$C_m(\epsilon_1, \epsilon_2) = C(\epsilon_1)C(\epsilon_2)$$

for $m = 1, 2$ and for all ϵ_1, ϵ_2 then $\alpha = 0$.

Proof: This follows by repeated use of the following identities for a $N(0,1)$ random variable z:

$$\begin{aligned} \mathsf{E}\left[e^{iuz}\right] &= e^{-\frac{1}{2}u^2} \\ \mathsf{E}\left[e^{-\frac{1}{2}u^2z^2}\right] &= (1+u^2)^{-\frac{1}{2}}. \end{aligned}$$

For the $\{x_t\}$ process,

$$\begin{aligned}
\mathsf{E}\left[\exp\left(iux_t + ivx_{t+1}\right)\right] &= \mathsf{E}\left[\exp\left(iu\epsilon_t + iu\alpha\epsilon_{t-1}\epsilon_{t-2} + iv\epsilon_{t+1} + iv\alpha\epsilon_t\epsilon_{t-1}\right)\right] \\
&= \exp\left(-\tfrac{1}{2}v^2\right)\mathsf{E}\left[\exp\left(iu\epsilon_t + iv\alpha\epsilon_t\epsilon_{t-1} + iu\alpha\epsilon_{t-1}\epsilon_{t-2}\right)\right] \\
&= \exp\left(-\tfrac{1}{2}v^2\right)\mathsf{E}\left[\exp\left(-\tfrac{1}{2}u^2\alpha^2\epsilon_{t-1}^2 + i(u + v\alpha\epsilon_{t-1})\epsilon_t\right)\right] \\
&= \exp\left(-\tfrac{1}{2}v^2\right)\mathsf{E}\left[\exp\left(-\tfrac{1}{2}u^2\alpha^2\epsilon_{t-1}^2 - \tfrac{1}{2}(u + v\alpha\epsilon_{t-1})^2\right)\right] \\
&= \exp\left(-\tfrac{1}{2}v^2 - \tfrac{1}{2}\tfrac{u^4}{u^2+v^2}\right) \\
&\qquad \mathsf{E}\left[\exp\left(-\tfrac{1}{2}\left(\tfrac{uv}{(u^2+v^2)^{\frac{1}{2}}} + \alpha(u^2+v^2)^{\frac{1}{2}}\epsilon_{t-1}\right)^2\right)\right] \\
&= \exp\left(-\tfrac{1}{2}v^2 - \tfrac{1}{2}\tfrac{u^4}{u^2+v^2}\right)\frac{\exp\left(-\frac{u^2v^2}{2(u^2+v^2)(1+\alpha^2(u^2+v^2))}\right)}{(1+\alpha^2(u^2+v^2))^{\frac{1}{2}}}
\end{aligned}$$

Denote this last expression by $\phi(u,v)$. Then

$$\begin{aligned}
\phi(0,v) &= \frac{e^{-\frac{1}{2}v^2}}{(1+\alpha^2v^2)^{\frac{1}{2}}} \\
\phi(u,0) &= \frac{e^{-\frac{1}{2}u^2}}{(1+\alpha^2u^2)^{\frac{1}{2}}}
\end{aligned}$$

and condition (2.2) holds if and only if $\alpha = 0$, i.e., the process is independent. □

Theorem 7: *If $K(\epsilon_i) - C(\epsilon_i)^2 > 0$ for $i = 1,2$ and if the random variables $\{X_t\}$ are independent, then*

$$\sqrt{n}\frac{S_{m,n}(\epsilon_1,\epsilon_2)}{\sigma(\epsilon_1,\epsilon_2)} \xrightarrow{D} N(0,1). \tag{A.4}$$

The asymptotic variance can be consistently estimated by:

$$\sigma_n^2(\epsilon_1,\epsilon_2) = 4[\hat{K}_n(\epsilon_1) - \hat{C}_n(\epsilon_1)^2][\hat{K}_n(\epsilon_2) - \hat{C}_n(\epsilon_2)^2]$$

where

$$\hat{K}_n(\epsilon) = \frac{1}{n^3}\sum_{r=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\chi_\epsilon(|x_r - x_s|)\chi_\epsilon(|x_s - x_t|) \tag{A.5}$$

and

$$\hat{C}_n(\epsilon) = \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\chi_\epsilon(|x_s - x_t|). \tag{A.6}$$

Proof: *This is a direct consequence of Denker and Keller (1983, Theorem 1, p507) applied to the kernel in equation (3.8).* □

Theorem 8: If $K(\epsilon) - C(\epsilon)^2 > 0$ then $\exists \alpha < \infty$

$$||F_n(\cdot;\epsilon) - \Phi||_\infty = \sup_t |F_n(t;\epsilon) - \Phi(t)| \leq \frac{\alpha C(\epsilon)}{8\sqrt{n}[K(\epsilon) - C(\epsilon)^2]^{\frac{3}{2}}}$$

where Φ is the standard Normal distribution function.

Proof: *The inequality comes from Callaert and Janssen (1978).*

Theorem 9: Let $\{X_t\}$ be IID. Then

$$\mathcal{P}(\lim_{n\to\infty} \sup_\epsilon |I_{1,n}(\epsilon) - C(\epsilon)| = 0) = 1.$$

Proof: The proof follows the same technique as the Glivenko–Cantelli Theorem[5] The proof for the case that F is continuous (in which case C is continuous) is presented here. Define $\{\epsilon_{ij}\}$ by:

$$\epsilon_{ij} = \inf\left\{\epsilon \mid \tfrac{i}{j} \leq C(\epsilon)\right\}$$

for $i = 0, \ldots, j$ and $j = 1, 2, \ldots$. Now let

$$A_{ij} = \left\{I_{1,n}(\epsilon_{ij}) \to C(\epsilon_{ij})\right\}$$

and $A = \bigcap_{j=1}^{\infty} \bigcap_{i=0}^{j} A_{ij}$. Since $\mathrm{P}(A_{ij}) = 1$ it follows that $\mathrm{P}(A) = 1$ as well. The event A is:

$$A = \{\lim_{j\to\infty} \max_{0\leq i\leq j} |I_{1,n}(\epsilon_{ij}) - C(\epsilon_{ij})| = 0\}.$$

For $\epsilon_{ij} \leq \epsilon < \epsilon_{i+1,j}$ we have

$$C(\epsilon_{ij}) \leq C(\epsilon) \leq C(\epsilon_{i+1,j}) \text{ and } I_{1,n}(\epsilon_{ij}) \leq I_{1,n}(\epsilon) \leq I_{1,n}(\epsilon_{i+1,j})$$

and by construction of the sequence $\{\epsilon_{ij}\}$,

$$0 \leq C(\epsilon_{i+1,j}) - C(\epsilon_{ij}) \leq \frac{1}{j}.$$

Therefore,

$$I_{1,n}(\epsilon) - C(\epsilon) \leq I_{1,n}(\epsilon_{i+1,j}) - C(\epsilon_{ij}) \leq I_{1,n}(\epsilon_{i+1,j}) - C(\epsilon_{i+1,j}) + \tfrac{1}{j}$$

and

$$I_{1,n}(\epsilon) - C(\epsilon) \geq I_{1,n}(\epsilon_{i,j}) - C(\epsilon_{i+1,j}) \geq I_{1,n}(\epsilon_{i,j}) - C(\epsilon_{i,j}) - \tfrac{1}{j}.$$

Hence for all ϵ and j

$$|I_{1,n}(\epsilon) - C(\epsilon)| \leq \max_{0\leq i\leq j} |I_{1,n}(\epsilon_{ij}) - C(\epsilon_{ij})| + \tfrac{1}{j}$$

and the conclusion follows. □

[5] See, fro example, Loève (1963, p.20).

References

BROCK, W. A. (1986): "Distinguishing Random and Deterministic Systems: Abridged Version," *Journal of Economic Theory*, 40, 168–195.

BROCK, W. A., and W. D. DECHERT (1988): "Theorems on Distinguishing Deterministic from Random Systems," in *Dynamic Econometric Modelling*, ed. by W. Barnett, E. Berndt, and H. White, pp. 247–265, Cambridge. Cambridge University Press.

BROCK, W. A., W. D. DECHERT, and J. A. SCHEINKMAN (1987): "A Test for Independence Based On the Correlation Dimension," Department of Economics, University of Wisconsin at Madison, University of Houston, and University of Chicago.

BROCK, W. A., W. D. DECHERT, J. A. SCHEINKMAN, and B. LEBARON (1996): "A Test For Independence Based on the Correlation Dimension," *Econometric Reviews*, 15(3), 197–235.

BROCK, W. A., D. HSIEH, and B. LEBARON (1991): *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*. MIT Press, Cambridge, Massachusetts.

CALLAERT, H., and P. JANSSEN (1978): "The Berry–Esseen Theorem for U–Statistics," *The Annals of Statistics*, 6(2), 417–421.

DENKER, M., and G. KELLER (1983): "On U-Statistics and von Mises Statistics for Weakly Dependent Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64, 505–522.

ENGLE, R. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987–1007.

GRASSBERGER, P., and I. PROCACCIA (1983a): "Estimating the Kolmogorov Entropy from a Chaotic Signal," *Physical Review*, 28A, 25–91.

——— (1983b): "Measuring the Strangeness of Strange Attractors," *Physica*, 9D, 189.

LOÈVE, M. (1963): *Probability Theory*. Van Nostrand, New York, 3rd edn.

POLLARD, D. (1984): *Convergence of Stochastic Processes*. Springer-Verlag, New York.

RAMSEY, J. B., and H.-J. YUAN (1987): "The Statistical Properties of Dimension Calculations Using Small Data Sets," New York University.

SEN, P. K. (1963): "On the Properties of U–Statistics When the Observations are not Independent," *Bulletin of the Calcutta Statistical Association*, 12, 69–92.

SERFLING, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

THEILER, J. (1988): "Statistical Precision of Dimension Estimators," *Physics Review*, 41A(6), 3038–3051.