PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Independence in Time Series: Another Look at the BDS Test [and Discussion]

Rodney C. L. Wolff and P. M. Robinson

Phil. Trans. R. Soc. Lond. A 1994 **348**, 383-395
doi: 10.1098/rsta.1994.0098

Independence in time series: another look at the BDS test†

By Rodney C. L. Wolff
Department of Statistics, University of Glasgow, Glasgow G12 8QW, U.K.

This paper examines a statistical method derived from chaos theory. The correlation integral was proposed over a decade ago as a way of detecting chaos in a possibly partial realization of a dynamical system, because it depends on the spatial arrangement of the reconstructed attractor of the system. We exploit geometrical properties of an embedded time series to establish a test of independence in the original time series. Earlier efforts here have used the Central Limit Theorem to obtain normality as the null distribution; however, the testing procedure was, to an extent, *ad hoc*. By making moderately weak assumptions about the marginal distribution of the given series, we obtain a Poisson law for the correlation integral under the null hypothesis of independence, and use non-parametric methods to specify the test precisely. We compare the size and power of the present test with its predecessor and with other non-parametric tests for serial dependence.

1. Introduction

(a) Motivation

In classical time series analysis, in which one uses linear models incorporating gaussian innovations, questions of independence are often dealt with by examining correlations. However, if there is departure from normality and linearity, then independence and lack of correlation need no longer be equivalent. A series with zero autocorrelation at each lag need not necessarily comprise independent observations (see, for example, Hall & Wolff 1994*b*). The existence of independence in a time series is of much interest in exploratory analyses and modelling, as well as in model-checking, since one typically fits models incorporating independent and identically distributed (IID) innovations (noise). Prediction and forecasting (Kendall & Ord 1990, p. 122) are also a concern in time series analysis. There exist many methods of forecasting which resort to using fitted parametric models, non-parametric approaches, and other *ad hoc* ideas, such as exploiting the 'momentum' of a series. In general, all such approaches are governed by dependence structures in the series, and precision of a forecast may be better in a correlated series than in another one of independent observations with identical marginal properties to those of the former. Hence, there is the need to examine dependence structures in time series without resorting to statistical correlations or other lin-

† This paper was produced from the author's disk by using the TEX typesetting system.

Phil. Trans. R. Soc. Lond. A (1994) **348**, 383–395
Printed in Great Britain

ear features. The correlation integral provides a useful medium. Moreover, the use of the correlation integral in assessing independence in a time series is a fine example of the use of chaos theory in the context of stochastic, non-chaotic processes.

(*b*) *The correlation integral*

We give a brief introduction to the correlation integral. An accompanying paper in this issue by Dr Cutler, as well as the comprehensive review articles by Berliner (1992) and Isham (1993), provide more details.

A deterministic dynamical system can be specified in terms of a map, f, operating on a manifold, M, such that $f: M \to M$. The system may operate either in discrete time, in which case f is applied iteratively on some initial conditions, or in continuous time, in which case f specifies a collection of differential equations which produce a flow beginning at the initial conditions. Suppose that f produces a discrete time trajectory, namely $\{x_0, f(x_0), f^2(x_0), \ldots\}$, where $x_0 \in M$ is the initial condition and f^k denotes the k-fold iterate of f. The trajectory may be regarded as a time series. The long-run behaviour of a dynamical system, its equilibrium distribution, is embodied in the attractor. Therefore, an investigation of a dynamical system, given only a realization of it, often commences with an investigation of the nature of its attractor, and especially its dimension. The attractor of a chaotic dynamical system can be Cantor-like, and so may be summarized using a fractional dimension.

We can define a fractional dimension as follows. Choose a point x_0, specified in d coordinates, from the attractor. If Δ is the dimension of the attractor, then, heuristically, the number of points, $N(h)$, of a 'typical' trajectory of length n of the dynamical system within distance h of x_0 should satisfy $N(h) \propto h^{\Delta}$. This leads to the definition of the information or correlation dimension,

$$\Delta \equiv \lim_{n\to\infty} \lim_{h\to 0^+} \frac{\log N(h)}{\log h}.$$

If the limit exists then it need not be an integer. The observed realization of a dynamical system may be a cross-section of the true realization, for example, a projection into a lower-dimensional space or the realization of the first-return map to a lower-dimensional subspace. Takens (1981) proves that such a cross-section of a generic dynamical system with attractor of dimension Δ can be embedded in a euclidean space of dimension at most $2\Delta + 1$. Grassberger & Procaccia (1983) provide an algorithm to obtain the embedding dimension, given a time series $Y_1, \ldots, Y_n$. They use the correlation integral,

$$C(n,p,h) \equiv \binom{n-p+1}{2}^{-1} \sum_{i=1}^{n-p} \sum_{j=i+1}^{n-p+1} I(||V_i - V_j|| \leqslant h), \tag{1.1}$$

where the vectors

$$V_i \equiv (Y_i, \ldots, Y_{i+p-1})' \qquad (i = 1, \ldots, n-p+1) \tag{1.2}$$

constitute the embedding of the time series into p-dimensional euclidean space, and where $I(A)$ denotes the indicator function of event A. Assuming that the underlying series is completely random, in the sense of having no deterministic structure, then the number of pairs of points which are within distance h of each other should be proportional to the volume containing such points, h^p. Complete spatial randomness in spatial point patterns is an example of this (Diggle 1983). If the series is deterministic or chaotic, then vectors of length p_0 or more, for some p_0, will contain functionally dependent entries, and so the power law will saturate at h^{p_0} when $p \geqslant p_0$. Suitable log–log plots can be used to check for saturation in the power law. Smith (1988) discusses limitations in this regard.

(*c*) *A test for independence (Brock et al. 1986)*

Brock *et al.* (1986) apply the Central Limit Theorem to the correlation integral and deduce its asymptotic distribution. Their definition of the correlation integral uses the L^∞ norm in equation (1.1) for comparing distances, h, between points of the embedded series. We conjecture that the choice of norm will not substantially affect the procedure, as $h \to 0^+$, if the underlying series is independent. Brock *et al.* prove that, under the null hypothesis that $Y_1, \ldots, Y_n$ constitute a series of IID observations, $C(n,p,h)$ converges to $\{\text{pr}(|Y_i - Y_j| \leqslant h)\}^p$ as $n \to \infty$, with probability 1. Further, they show that $\alpha_1 C(n,p_1,h) + \alpha_2 C(n,p_2,h)$, suitably scaled, has an asymptotic normal distribution, the variance of which is given explicitly in terms of $\text{pr}(|Y_i - Y_j| \leqslant h)$ and $\text{pr}(|Y_i - Y_j| \leqslant h \;\&\; |Y_j - Y_k| \leqslant h)$ $(i, j, k$ distinct). There is a numerical study, which we shall discuss more fully along with our results, in which they observe that the actual type I error exceeds the nominal significance when $p \geqslant 3$. Type II error is calculated for three specific parametric alternative hypotheses.

Brock *et al.* do not place any conditions on the order of magnitude of the comparison distance, h. If h becomes large then the correlation integral equals unity with probability 1: it is obvious that this can occur for finite h. For fixed n, as $h \to 0^+$ there will be fewer and fewer points of the embedded series deemed sufficiently close and therefore $C(n,p,h) \to 0$. We suspect that their simulated type I error would converge to the nominal significance level if h were chosen in accordance with their rate of convergence in distribution to the normal of the test statistic.

(*d*) *Some other tests for independence in a time series*

Skaug & Tjøstheim (1993*a*) consider estimating the dependence measure proposed by Hoeffding (1948) for the case of two random variables, which consists of calculating the squared error between the bivariate cumulative distribution function (CDF) and the product of the marginals. Skaug & Tjøstheim consider the time series case by estimating Hoeffding's measure via the empirical bivariate CDF The bivariate CDF for a time series is defined to be that of observations at successive epochs. In particular, they obtain the empirical marginal CDFs directly from the bivariate empirical CDF, and they note the relationship of their test statistic with the Cramér–von Mises functional for test of fit. Skaug & Tjøstheim (1993*b*) similarly construct a test using kernel estimates of the bivariate and marginal densities.

Some classical tests for independence are based on the distributions of partial autocorrelations, as described in Hannan (1970, ch. 3). The Durbin–Watson statistic (Durbin & Watson 1950, 1951, 1971) enjoys popularity in testing for serial correlation.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

(e) Outline of paper

The paper is organized as follows. In § 2 *a* we outline the Chen–Stein method for proving Poisson convergence of a sum of correlated binary random variables. In § 2 *b*, we apply this to the (unscaled) correlation integral, as in (1.1), and, in particular, show that convergence relies only on a technical condition on the marginal distribution of the observed time series. The proof further yields a relationship which governs the magnitude of the comparison distance, namely $h \propto n^{-2/p}$, in the notation of (1.1). In § 3 *a*, we obtain the mean of the test statistic under the null hypothesis of independence in terms of integrals of second, third and fourth powers of the marginal density of the given time series. Power considerations are discussed in § 3 *b*. A numerical study in § 4 illustrates some size and power features of the present test, and makes contrasts with related non-parametric tests for independence. Some persisting difficulties are summarized in § 5.

2. Poisson convergence

(a) The Chen–Stein method

Chen (1975), using methods established by Stein (1972), proved that convergence to a Poisson distribution, for the number of occurrences of dependent events, can often be deduced by computing no more than the first and second moments. Arratia *et al.* (1989, 1990) review this so-called Chen–Stein method and discuss extensions to the multivariate case and to approximations by Poisson processes, along with applications (see also Barbour *et al.* 1992). We shall use the following result.

Let $\mathcal{I}$ be an index set and let X_k $(k \in \mathcal{I})$ be Bernoulli random variables with $0 \leqslant p_k = \mathrm{pr}(X_k = 1) = E(X_k) < 1$. Let

$$D \equiv \sum_{k\in\mathcal{I}} X_k, \qquad \lambda \equiv E(D) = \sum_{k\in\mathcal{I}} p_k,$$

such that $0 < \lambda < \infty$. For each $k \in \mathcal{I}$, define

$$\mathcal{I}_k \equiv \mathcal{I} \backslash \{\ell\, ;\ X_\ell \text{ is independent of } X_k\},$$

which Arratia *et al.* call a neighbourhood of dependence. Put

$$b_1 \equiv \sum_{k\in\mathcal{I}} \sum_{\ell\in\mathcal{I}_k} p_k p_\ell, \quad b_2 \equiv \sum_{k\in\mathcal{I}} \sum_{\ell\in\mathcal{I}_k\backslash\{k\}} p_{k\ell},$$

$$b_3 \equiv \sum_{k\in\mathcal{I}} E|E\{X_k - p_k \mid \sigma(X_\ell\, ;\ \ell \in \mathcal{I}\backslash\mathcal{I}_k)\}|,$$

where $p_{k\ell} = E(X_k X_\ell)$ and σ denotes the sigma field generator. All of b_1, b_2 and b_3 are non-negative. Let P have a Poisson distribution such that $E(P) = E(D)$, and let

$$\textsc{tvd}(P, D) \equiv 2 \sup_{A\subseteq \boldsymbol{Z}} |\mathrm{pr}(P \in A) - \mathrm{pr}(D \in A)|$$

denote the total variation distance between the distributions of P and D. Arratia *et al.* (1989) show that $\textsc{tvd}(P, D) \leqslant 2(b_1 + b_2 + b_3)$.

(b) *Proof of Poisson convergence of the correlation integral*

Berman & Eagleson (1983) obtain Poisson limit results for a family of incomplete symmetric statistics, of which the correlation integral is a member. We sketch the proof here, as some details will provide insight for construction of the test statistic.

Let the time series $Y_1, \ldots, Y_n$ be IID with a bounded, continuous probability density function (PDF) g. With $n' = n - p + 1$, define an embedding of the original time series in a phase space of dimension p, as in (1.2). The binary-valued variables $X_{ij} = I\{V_i - V_j \in \mathcal{B}_p(0,h)\}$ $(1 \leqslant i < j \leqslant n')$ will play the roles of the Bernoulli variables in the Chen–Stein theorems. Note that $\mathcal{B}_p(0,h)$ denotes the ball of radius h in p-dimensional euclidean space. Define the correlation count to be

$$D(n,p,h) \equiv \sum_{i=1}^{n'-1} \sum_{j=i+1}^{n'} X_{ij} = \sum_{i=1}^{n'-1} \sum_{j=1}^{p-1} X_{i,i+j} + \sum_{i=1}^{n'-1} \sum_{j=p}^{n'-i} X_{i,i+j}.$$

Clearly, it is $n'(n'-1)/2$ times the value of the correlation integral, given in (1.1). Let $\mathcal{S}_j = \{X_{i,i+j}\}_{i=1}^{n'-i}$ be a sequence of $M_{p,j}$-dependent random variables, such that $M_{p,j} = (p-j)I(1 \leqslant j \leqslant p-1)$, and where the sequence $\{U_i\}$ of random variables is said to be M-dependent if U_i is independent of $\sigma(U_{i+M+1}, U_{i+M+2}, \ldots)$ $(i = 1, 2, \ldots)$. When $M = 0$, M-dependence corresponds to independence.

For $i < j$,

$$E(X_{ij}) = \text{pr}(X_{ij} = 1) = \text{pr}\{V_i - V_j \in \mathcal{B}_p(0,h)\} \sim h^p v_p g_{ij}(0) \qquad (2.1)$$

as $h \to 0^+$, where v_p is the volume of $\mathcal{B}_p(0,1)$ and g_{ij} is the PDF of $V_i - V_j$. Hence the expected contribution to the value of $D(n,p,h)$ by the sequences, $\mathcal{S}_j$, of $M_{p,j}$-dependent random variables $(j = 1, \ldots, p-1)$ is of order of magnitude nh^p as $n \to \infty$ and $h \to 0^+$. Also, the expected contribution to the value of $D(n,p,h)$ by the sequences, $\mathcal{S}_j$, of independent random variables $(j = p, \ldots, n'-i)$ is of order of magnitude n^2h^p under the same limits.

Recall that a Binomial random variable with index N and parameter π converges in distribution to a Poisson random variable with parameter λ as $N \to \infty$ provided only that $N\pi \to \lambda$ as well. Along the same lines as this, the preceding paragraph suggests that, provided n^2h^p tends to a constant, say L, as $n \to \infty$ and $h \to 0^+$, $D(n,p,h)$ will be asymptotically bounded in expectation, and Poisson convergence may be possible. Indicate this multiple limit with the notation $(n, h, n^2h^p) \to (\infty, 0^+, L)$. Specifically, if $0 \leqslant \sup E\{D(n,p,h)\} \leqslant \beta < \infty$, where the supremum is taken over n, then $\text{pr}\{D(n,p,h) > d\} \leqslant E\{D(n,p,h)\}/d \leqslant \beta/d$, by Markov's inequality, as $D(n,p,h) \geqslant 0$ with probability 1. Hence the probability mass of $D(n,p,h)$ is concentrated towards the origin, and this is consistent with $D(n,p,h)$ attaining a Poisson distribution in the limit. Modifying the notation of Arratia *et al.* (1989) in the obvious way, let $\mathcal{I} = \{(i,j); 1 \leqslant i < j \leqslant n'\}$, $\mathcal{I}_{ij} = \{(k,\ell)\ ;\ X_{ij}$ is dependent upon $X_{k\ell}\}$, and so $(i,j) \in \mathcal{I}_{ij}$. Since X_{ij} is dependent on $X_{k\ell}$ exactly when $i-p+1 \leqslant k \leqslant i+p-1$ or $j-p+1 \leqslant \ell \leqslant j+p-1$, $\#(\mathcal{I}_{ij}) \leqslant (2p-1)^2$ and $\#(\mathcal{I}) = n'(n'-1)/2$. Thus

$$b_1 \leqslant \#(\mathcal{I})\,\#(\mathcal{I}_{ij}) \left\{h^p v_p \max_{(i,j)\in\mathcal{I}} g_{ij}(0)\right\}^2 \leqslant \binom{n'}{2} (2p-1)^2 h^{2p} v_p^2 \max_{(i,j)\in\mathcal{I}} \{g_{ij}(0)\}^2.$$

Since g is bounded so too is g_{ij}, and so $b_1 \to 0$ as $(n, h, n^2h^p) \to (\infty, 0^+, L)$.

Now consider b_2. Choose $(i,j) \in \mathcal{I}$ and $(k,\ell) \in \mathcal{I}_{ij}$. Since $i < j$ and $k < \ell$ there is an index $s \in S_1 = \{i,\ldots,i+p-1\} \cup \{j,\ldots,j+p-1\}$ such that $s \notin S_2 = \{k,\ldots,k+p-1\} \cup \{\ell,\ldots,\ell+p-1\}$. Let $\mathcal{X} = \{X_t\ ;\ t \in S_1 \cup S_2\} \backslash \{X_s\}$. Then

$$\begin{aligned}
E\,(X_{ij}X_{k\ell}) &= \mathrm{pr}\{V_i - V_j \in \mathcal{B}_p(0,h)\ \&\ V_k - V_\ell \in \mathcal{B}_p(0,h)\} \\
&= \mathrm{pr}\{V_k - V_\ell \in \mathcal{B}_p(0,h) \mid V_i - V_j \in \mathcal{B}_p(0,h)\}\mathrm{pr}\{V_i - V_j \in \mathcal{B}_p(0,h)\} \\
&\leqslant \mathrm{pr}\{V_k - V_\ell \in \mathcal{B}_p(0,h) \mid \mathcal{X}\}\mathrm{pr}\{V_i - V_j \in \mathcal{B}_p(0,h)\} \\
&= \mathrm{pr}(|X_s - x_0| \leqslant h \mid \mathcal{X})\mathrm{pr}\{V_i - V_j \in \mathcal{B}_p(0,h)\} \\
&= \left\{\int_{-h}^{h} g_X(x - x_0)\,\mathrm{d}x\right\}\mathrm{pr}\{V_i - V_j \in \mathcal{B}_p(0,h)\} \\
&\sim 2h\left\{\sup_u g_X(u)\right\} h^p v_p g_{ij}(0),
\end{aligned}$$

as $h \to 0^+$, where g_X is the PDF of X_s. Thus b_2 is dominated by a quantity of order $n^2 2h\,\{\sup_u g_X(u)\}\, h^p v_p g_{ij}(0) \to 0$ as $(n,h,n^2h^p) \to (\infty, 0^+, L)$.

By construction of $\mathcal{I}_{ij}$, $b_3 = 0$.

Hence we have that $D(n,p,h)$ has a limiting Poisson distribution as described. In practice, the choice of the comparison distance, h, will be governed by the relationship $h \propto n^{-2/p}$.

3. Testing procedure

(a) Calculation of the mean in the Poisson law

Under the prescribed rate of convergence, the asymptotic mean of the correlation count is $\frac{1}{2}Lv_p\bar{g}_0$, where $\bar{g}_0$ denotes the average of the densities of each $V_i - V_j$ $(i < j)$.

Since $v_p = \pi^{p/2}/\Gamma\{\frac{1}{2}(p+2)\}$ (Courant 1937, p. 304), it remains to evaluate the density of $V_i - V_j$ at zero, which is a function of the PDF of $Y_i - Y_j$. Let g^* be the PDF of $Y_1 - Y_2$. Then

$$g^*(u_1) = \int_{\mathcal{R}} g(u_1 + u_2)g(u_2)\,\mathrm{d}u_2, \qquad g^*(0) = \int_{\mathcal{R}} \{g(u)\}^2\,\mathrm{d}u.$$

Observe that V_i and V_j $(i < j)$ are stochastically independent whenever $j - i \geqslant p$. Suppose this to be the case. Then, for infinitesimals $\mathrm{d}w_0,\ldots,\mathrm{d}w_{p-1}$, the PDF, $g_{ij,1}$, of $V_i - V_j$, evaluated at zero, is given by

$$\begin{aligned}
g_{ij,1}(0)\mathrm{d}w_0 \cdots \mathrm{d}w_{p-1} &= \mathrm{pr}(0 < Y_{i+k} - Y_{j+k} \leqslant \mathrm{d}w_k,\ 0 \leqslant k \leqslant p-1) \\
&= \prod_{k=1}^{p-1} \mathrm{pr}(0 < Y_{i+k} - Y_{j+k} \leqslant \mathrm{d}w_{k+1}),
\end{aligned}$$

since the entries of $V_i - V_j$ are mutually independent, which yields

$$g_{ij,1}(0) = \{g^*(0)\}^p = \left[\int_{\mathcal{R}} \{g(y)\}^2\,\mathrm{d}y\right]^p. \tag{3.1}$$

Consider the case $\lfloor\frac{1}{2}(p+1)\rfloor \leqslant j - i < p$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding x. Let $\mathcal{Y}_{ij} = \{Y_k; Y_k$ is an entry in both V_i and $V_j, 1 \leqslant k \leqslant n\}$. Under the condition $j - i = m$, say, no two elements of $\mathcal{Y}_{ij}$ will occupy a corresponding entry of both V_i and V_j. In particular, precisely the first $p-m$ entries of V_i and the last $p-m$ entries of V_j are contained in $\mathcal{Y}_{ij}$. It is important to note this in what follows, as we shall condition on $\mathcal{Y}_{ij}$. It is easily checked that $2(p-m) \leqslant p$. Now, let $V_i - V_j$ have PDF $g_{ij,2}$. Then, for infinitesimals $\mathrm{d}w_0, \ldots, \mathrm{d}w_{p-1}$,

$$\begin{aligned} g_{ij,2}(0)\mathrm{d}w_0 \cdots \mathrm{d}w_{p-1} &= \mathrm{pr}(0 < Y_{i+k} - Y_{j+k} \leqslant \mathrm{d}w_k,\ 0 \leqslant k \leqslant p-1) \\ &= \int \cdots \int \mathrm{pr}(0 < Y_{i+k} - Y_{j+k} \leqslant \mathrm{d}w_k,\ 0 \leqslant k \leqslant p-1 \\ &\qquad |y_r < Y_r \leqslant y_r + \mathrm{d}y_r,\ Y_r \in \mathcal{Y}_{ij}) \times \left\{ \prod_{\ell=m}^{p-1} g(y_{i+\ell})\,\mathrm{d}y_{i+\ell} \right\}, \end{aligned}$$

since $\mathrm{pr}(y_r < Y_r \leqslant y_r + \mathrm{d}y_r) = g(y_r)\,\mathrm{d}y_r$ for infinitesimal $\mathrm{d}y_r$, and where integration is over $y_{i+m} \in \mathcal{R}, \ldots, y_{i+p-1} \in \mathcal{R}$. So, as $\mathrm{d}w_k \to 0$ $(k = 0, \ldots, p-1)$, by continuity of g,

$$g_{ij,2}(0) = \{g^*(0)\}^{2m-p} \int \cdots \int \left\{ \prod_{k=m}^{p-1} g(y_{i+k}) g(y_{i+k}) \right\} \left\{ \prod_{\ell=m}^{p-1} g(y_{i+\ell})\,\mathrm{d}y_{i+\ell} \right\} \quad (3.2)$$

$$\begin{aligned} &= \{g^*(0)\}^{2m-p} \left[\int_{\mathcal{R}} \{g(y)\}^3\,\mathrm{d}y \right]^{2p-2m} \\ &= \left[\int_{\mathcal{R}} \{g(y)\}^2\,\mathrm{d}y \right]^{2m-p} \left[\int_{\mathcal{R}} \{g(y)\}^3\,\mathrm{d}y \right]^{2p-2m}. \end{aligned} \quad (3.3)$$

We make two remarks. First, the product of squares of $g(y_k)$ in equation (3.2) arises from the first $(i+p-1)-(i+m)+1 = p-m$ entries of $V_i - V_j$, and the last $p-m$ of same, since, in those entries, we condition on exactly one term of the difference $Y_{i+k} - Y_{j+k}$. The remaining factors in equation (3.2) contribute to the integral of cubes of g in (3.3). Secondly, when $m = p$ we obtain the result of the first case, namely when $j - i \geqslant p$.

Finally, let $1 \leqslant j - i < \lfloor \frac{1}{2}(p+1) \rfloor$ $(p > 1)$, and define $\mathcal{Y}_{ij}$ as in the case $\lfloor \frac{1}{2}(p+1) \rfloor \leqslant j - i < p$. Then the number of entries in $V_i - V_j$ which comprise a difference of two elements of $\mathcal{Y}_{ij}$ is $2(p-m) - p = p - 2m$, since precisely the first $p-m$ entries of V_i and the last $p-m$ entries of V_j are elements of $\mathcal{Y}_{ij}$, and $V_i - V_j$ is of length p. So the PDF of $V_i - V_j$ at zero is given by

$$\begin{aligned} g_{ij,3}(0)\,\mathrm{d}w_0 \cdots \mathrm{d}w_{p-1} &= \mathrm{pr}(0 < Y_{i+k} - Y_{j+k} \leqslant \mathrm{d}w_k,\ 0 \leqslant k \leqslant p-1) \\ &= \int \cdots \int \mathrm{pr}(0 < Y_{i+k} - Y_{j+k} \leqslant \mathrm{d}w_k,\ 0 \leqslant k \leqslant p-1 \\ &\qquad |y_r < Y_r \leqslant y_r + \mathrm{d}y_r,\ Y_r \in \mathcal{Y}_{ij}) \times \left\{ \prod_{\ell=m}^{p-1} g(y_{i+\ell})\,\mathrm{d}y_{i+\ell} \right\} \\ &= \int \cdots \int \mathrm{pr}(0 < Y_{i+k} - y_{j+k} \leqslant \mathrm{d}w_k,\ 0 \leqslant k \leqslant m-1\ \& \\ &\qquad 0 < y_{i+\ell} - Y_{j+\ell} \leqslant \mathrm{d}w_\ell,\ p-m \leqslant \ell \leqslant p-1\ \& \\ &\qquad 0 < y_{i+r} - y_{j+r} \leqslant \mathrm{d}w_r,\ m \leqslant r \leqslant p-m-1) \\ &\qquad \times \left\{ \prod_{s=m}^{p-1} g(y_{i+s})\,\mathrm{d}y_{i+s} \right\}, \end{aligned}$$

where integration is over $y_{i+m} \in \mathcal{R}, \ldots, y_{i+p-1} \in \mathcal{R}$. As

$$\mathrm{d}w_k \to 0 (k = 0, \ldots, m-1, p-m, \ldots, p-1),$$

it can be shown that

$$g_{ij,3}(0)\,\mathrm{d}w_m \cdots \mathrm{d}w_{p-m-1} = \int \cdots \int \left[\prod_{k=m}^{2m-1} \{g(y_{i+k})\}^2\,\mathrm{d}y_{i+k}\right] \left\{\prod_{\ell=p-m}^{p-1} g(y_{i+\ell})\right\}$$
$$\times \left\{\prod_{s=2m}^{p-1} g(y_{i+s})\,\mathrm{d}y_{i+s}\right\} \left\{\prod_{r=m}^{p-m-1} I\left(0 < y_{i+r} - y_{i+m+r} \leqslant \mathrm{d}w_r\right)\right\}. \quad (3.4)$$

This leads to two subcases.

The first subcase is when $2m-1 < p-m$, or $p \geqslant 3m$. This is possible, for example if $p = 4$ and $m = 1$. The presence of the indicator functions in (3.4) restricts the integration subject to $y_{i+2m} = y_{i+m}, \ldots, y_{i+p-m} = y_{i+p-2m}, \ldots, y_{i+p-1} = y_{i+p-1-m}$, a total of $2p - m$ constraints, as $\mathrm{d}w_k \to 0$ $(k = m, \ldots, p-m-1)$. It may be shown that

$$g_{ij,3}(0) = \left[\int_{\mathcal{R}} \{g(y)\}^3\,\mathrm{d}y\right]^{p-3m} \left[\int_{\mathcal{R}} \{g(y)\}^4\,\mathrm{d}y\right]^{4m-p} \left[\int_{\mathcal{R}} \{g(y)\}^2\,\mathrm{d}y\right]^{p-3m}. \quad (3.5)$$

The second subcase is when $p < 3m$. This is possible, for example if $p = 5$ and $m = 2$. The calculations become quite complicated in this case, although it appears that we recover an expression very similar to that in equation (3.5).

Observe that the number of cases where $j - i \leqslant p$ is $o(1)$ asymptotically.

It is preferable either to calculate or to estimate the integrals of various powers of g directly, as in Hall & Marron (1987). However, theory in that and related papers deals only with integrals of squares of density derivatives. Motivated by this particular application, Hall & Wolff (1994*a*) have developed consistent estimators of integrals of all powers of density derivatives, and we shall use their calculations in the sequel.

(*b*) *A sequence of statistical hypotheses*

The result of § 2 *b*, namely that, under the stated conditions, the correlation count is asymptotically distributed according to a Poisson distribution, enables us to construct a test of the hypothesis

$$H_0 : Y_1, \ldots, Y_n \text{ are IID}$$

against an alternative hypothesis specified by some nonlinear model for the series $Y_1, \ldots, Y_n$, as in Brock *et al.* (1986).

As discussed in Cox & Hinkley (1974, § 4.8), the general unavailability of uniformly most powerful tests leads to the consideration of local alternatives: loosely, small departures from the null hypothesis.

For power considerations in the present problem, we may consider a particular family of local alternatives, defined by

$$H_q : Y_i = \sum_{j=0}^{q} Z_{i+j} \quad (i = 1, \ldots, n),$$

where $Z_1, \ldots, Z_{n+q}$ are IID. Moreover, H_q indeed specifies H_0 when $q = 0$.

If we assume that Z_1 possesses a bounded, absolutely continuous PDF, it is not

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

difficult to show that $D(n,p,h)$ will converge in law to a Poisson distribution as $n \to \infty$ and $h \to 0^+$ provided that, simultaneously, n^2h^p converges to a constant. Just as under H_0, we need only determine the density of $V_i - V_j$, evaluated at zero, to specify the limiting distribution. In principle this can be done in the same fashion as in § 3 *a*, but the expression will be cumbersome and tedious to evaluate. We describe a special case below, and study some more cases in § 4.

Suppose that $Z_1, \ldots, Z_{n+m} \sim N(0, \tau^2)$ and are, of course, independent. Let $q \geqslant 1$. Under H_q, $Y_i \sim N(0, q\tau^2)$ and $\mathrm{cov}(Y_i, Y_j) = (j-i)\tau^2 I(1 \leqslant m < q)$, where $m = j - i$, as before. Thus $V_i - V_j = A(Z_i, \ldots, Z_{j+p-1+q})'$ where A is a matrix of order $p \times (m+p+q)$ with the row arrangement

$$\underbrace{1,\ldots,1}_{a},\underbrace{0,\ldots,0}_{b},\underbrace{1,\ldots,1}_{a}$$

beginning in the (k,k) position $(k = 1, \ldots, p)$, and 0's everywhere else. Here, $a = q$ and $b = m - q$ if $m \geqslant q$, while $a = q - m$ and $b = m$ if $1 \leqslant m < q$. Consequently, $V_i - V_j$ has a multivariate normal distribution with mean 0 and variance matrix $\tau^2 AA'$, and the value of its density at zero is $\{(2\pi\tau^2)^p \det(AA')\}^{-1/2}$. With regard to hypothesis testing, this information now enables us to specify the distribution of the test statistic under any of the local alternative hypotheses, assuming normality of the components of the moving average representation of Y_i.

4. Numerical study

In all of the simulations, parameters were chosen or estimated subject to the mean of the test statistic being unity. Of course, this implies that the test statistic has a larger coefficient of variation than another with a larger mean. In the following results, this may be the cause of large variability, which is occasionally non-systematic. For example, convergence to unity of the power of the test against particular alternatives, illustrated below, can be better achieved with a larger mean for the test statistic.

Our test statistic is a discrete random variable, and a test with exact significance of 5% is not available. In all of the simulations, we used appropriate randomized rejection. That is, if c is such that $0 < \mathrm{pr}(D > c) = \theta < 0.05$ and $0 < \mathrm{pr}(D = c) = \theta'$ $(\theta + \theta' > 0.05)$, then reject the null hypothesis of independence if $D > c$, and reject the null hypothesis with probability $(0.05 - \theta)/\theta'$ if $D = c$; if $D < c$ then retain the null hypothesis in preference to the alternative.

Table 1 shows the probability of a Type I error when the given time series comprises IID uniform random variables on the interval [0,1]. The mean of the null Poisson distribution was calculated exactly, using the correct functional form of the respective densities. The sizes of the various simulated tests are roughly in accordance with the specified theoretical significance level of 5%. However, over the range of phase dimensions we considered, for each series length, the size increased slightly over the range $p = 2, \ldots, 5$ or so, and then fell, a curious result. This is discussed in § 5*b*.

Table 2 shows the probability of a Type I error when the given independent time series comprises (i) IID uniform random variables on the interval [0,1], and (ii) IID standard normal random variables. Randomized rejection was used, and the parameters of the respective test statistics were estimated non-parametrically. While larger sample sizes improve accuracy, the size of the test increases faster

Table 1. *Probability of a Type I error (IID uniform data): exact calculations*

(The table is headed by the dimension of the phase space. The remaining rows show the simulated probability of rejecting the null hypothesis that the given data are an IID time series, for independent time series of lengths 100, 250 and 1000. The parameter of the null Poisson distribution was evaluated according to equations (3.1), (3.3) and (3.5). There were 500 independent replications for each sample size.)

p	2	3	4	5	6	7	8
$n = 100$	0.044	0.042	0.042	0.046	0.056	0.040	0.028
$n = 250$	0.044	0.046	0.056	0.054	0.052	0.046	0.042
$n = 1000$	0.046	0.060	0.062	0.058	0.054	0.042	0.032

Table 2. *Probability of a Type I error (IID uniform and normal data): non-parametric approximations*

(The table is headed by the dimension of the phase space. The remaining rows show the simulated probability of rejecting the null hypothesis that the given data are an IID time series, for independent time series of lengths 100, 250 and 1000. The parameters of the null Poisson distribution were evaluated according to the non-parametric method described in the text. There were 1000 independent replications for each sample size.)

	p	2	3	4	5	6	7	8
uniform	$n = 100$	0.059	0.087	0.126	0.142	0.152	0.167	0.182
	$n = 250$	0.064	0.080	0.101	0.138	0.129	0.157	0.179
	$n = 1000$	0.053	0.082	0.096	0.119	0.112	0.136	0.164
normal	$n = 100$	0.056	0.075	0.093	0.117	0.133	0.146	0.209
	$n = 250$	0.066	0.076	0.089	0.102	0.113	0.143	0.172
	$n = 1000$	0.051	0.056	0.059	0.072	0.086	0.106	0.120

with phase dimension than it does in the case of non-randomized rejection, the results of which are omitted here for brevity. The results are especially pleasing in the normal case, from which one may conjecture that the test performs well on long-tailed data.

Figure 1 illustrates the power of the test for two particular alternatives, plotted against the parameter, α, where the models governing the alternative hypothesis are a bilinear (BL) process given by $X_t = (\alpha + \beta\epsilon_{t-1})X_{t-1} + \epsilon_t$ $(\alpha^2 + \beta^2 < 1)$ (here, $\beta = 0.4$), and a nonlinear moving average (NLMA) process given by $X_t = \epsilon_{t-1}(\alpha + \epsilon_t)$. These are two models considered in Skaug & Tjøstheim (1993*a*) and a related preprint. The embedding dimension is 2. The curve labelled I_0 refers to the present test; the curves labelled I_1 and I_2 correspond to Skaug & Tjøstheim's kernel-based statistic; the curve labelled I_3 corresponds to their statistic based on the empirical distribution function; and I_4 corresponds to a correlation measure, similar to the classical Durbin–Watson statistic, as in Durbin & Watson (1950, 1951, 1971). Broadly speaking, it would appear from these and other simulations that the present test compares well with other methods, except possibly in the case of some linear models where dependence is fairly weak. While its power improves as the embedding dimension increases, the size also increases, a feature

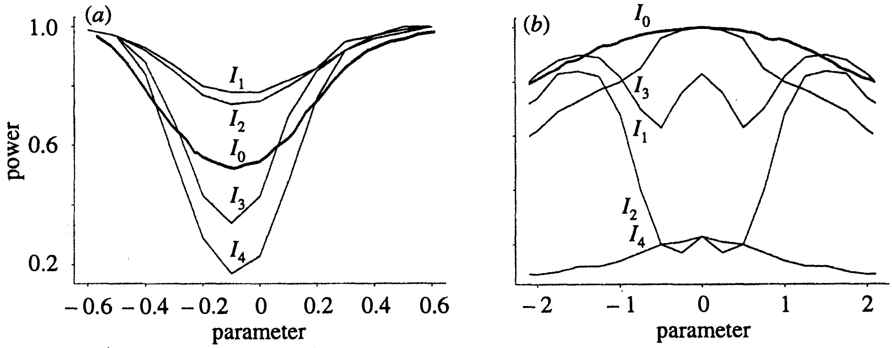

Figure 1. Graphs of power against the parameter, α, of the process governing the alternative hypothesis, namely (a) a bilinear process and (b) a nonlinear moving average process, as specified in §4. The power curve, I_0, relates to the present test; curves $I_1, \ldots, I_4$ relate to the non-parametric tests considered by Skaug & Tjøstheim (1993a), as mentioned in the text, and have been transcribed from their paper and a related preprint.

observed by Brock *et al.* (1986). Our simulations improved on the reported values of size and power from their numerical study.

5. Remarks

(a) *Present results*

We have shown that the unnormalized correlation integral may have a Poisson distribution in the presence of an arbitrary time series, and we have indicated how the comparison distance should be chosen to enable this.

Our distributional results improve on those of Brock *et al.* (1986) in two ways. First, we specify a suitable comparison distance for the correlation integral. Secondly, we improve slightly on simulated sizes of some tests.

One appealing property of our method is that it does not involve much more numerical computation than in the method of Skaug & Tjøstheim (1993a). Furthermore, the nature of the test statistic proposed by Skaug & Tjøstheim is comparable with $D(n,2,h)$: they examine the two-dimensional embedding of the given independent time series, particularly the departure of its bivariate density from the product of the two marginals. They also perform a numerical study of higher dimensional generalizations of their statistic. However, it is not clear that comparing a k-variate density with the product of its k marginals ($k > 2$) must necessarily account for all possible structures of serial dependence up to lag k. The correlation integral has the potential to do this, by computing clustering properties of k-dimensional vectors; of course, the curse of dimensionality is a substantial hindrance in this regard.

(b) *Further considerations*

The non-monotonicity of the simulated significance levels, as the embedding dimension increased, was perhaps due to large variability in the correlation count, another phenomenon which should be investigated further. It is believed that there would be further improvements if the mean of our test statistic were to be chosen optimally, in the sense of minimizing the coefficient of variation of the test statistic or seeking a variance–bias trade-off.

To compensate for the increase in the size of the test as the embedding dimension increases, it would be helpful to quantify the effect of the exponential growth of the phase space 'emptiness' on variation in the correlation count, with a view to introducing a correction factor to the comparison distance. Silverman (1986, ch. 4) discusses related issues in the context of multivariate density estimation.

An unresolved issue is the choice of embedding dimension for our test. Using the geometric analogue of Skaug & Tjøstheim's (1993*a*, *b*) methods, we need only consider two dimensions, for which the asymptotic distribution theory appears to apply in the finite sample case. On one hand, using only a two-dimensional embedding would appear to neglect autocorrelations at lags three or more; similarly for other embeddings. Nevertheless, it is the total geometric structure of the embedding of which the correlation integral is a measure. Given that there is no saturation in the power law for the correlation integral, higher-dimensional embeddings of independent data should contain neither any more nor any less structure than in two dimensions. However, it is not clear how to assess which aspects of dependence, other than possibly autocorrelations, are overlooked by an embedding dimension which is too small.

It is a pleasure to record my gratitude to Sir David Cox for supervision of this research, which comprises part of a recently completed doctoral thesis. This work was supported by an Australian Research Council grant to the Centre for Mathematics and its Applications, Australian National University, a computing equipment grant from the Wingate Foundation, London, and a travel grant from the Nuffield Foundation, London.

References

Arratia, R., Goldstein, L. & Gordon, L. 1989 Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Prob.* **17**, 9–25.

Arratia, R., Goldstein, L. & Gordon, L. 1990 Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5**, 403–434.

Barbour, A. D., Holst, L. & Janson, S. 1992 *Poisson approximation.* Oxford: Clarendon Press.

Berliner, L. M. 1992 Statistics, probability and chaos (with Discussion). *Statist. Sci.* **7**, 69–90.

Berman, M. & Eagleson, G. K. 1983 A Poisson limit theorem for incomplete symmetric statistics. *J. appl. Prob.* **20**, 47–60.

Brock, W. A., Dechert, W. D. & Scheinkman, J. A. 1986 A test for independence based on the correlation integral. Technical report: University of Wisconsin–Madison.

Chen, L. H. Y. 1975 Poisson approximation for dependent trials. *Ann. Prob.* **33**, 69–74.

Courant, R. 1937 *Differential and integral calculus*, vol. II. London: Blackie.

Cox, D. R. & Hinkley, D. V. 1974 *Theoretical statistics.* London: Chapman & Hall.

Diggle, P. J. 1983 *Statistical analysis of spatial point patterns.* London: Academic.

Durbin, J. & Watson, G. S. 1950 Testing for serial correlation in least-squares regression. *Biometrika* **37**, 409–428.

Durbin, J. & Watson, G. S. 1951 Testing for serial correlation in least-squares regression. *Biometrika* **38**, 159–178.

Durbin, J. & Watson, G. S. 1971 Testing for serial correlation in least-squares regression. *Biometrika* **58**, 1–19.

Grassberger, P. & Procaccia, I. 1983 Characterization of strange attractors. *Phys. Rev. Lett.* **50**, 346–349.

Hall, P. & Marron, J. S. 1987 Estimation of integrated squared density derivatives. *Statist. Prob. Lett.* **6**, 109–115.

Hall, P. & Wolff, R. C. L. 1994*a* Estimators of integrals of powers of density derivatives. *Statist. Prob. Lett.* (In the press.)

Hall, P. & Wolff, R. C. L. 1994*b* On the strength of dependence of a time series generated by a chaotic map. *J. Time Ser. Analysis.* (In the press.)

Hannan, E. J. 1970 *Multiple time series.* New York: Wiley.

Hoeffding, W. 1948 A nonparametric test of independence. *Ann. Math. Statist.* **19**, 546–557.

Isham, V. 1993 Some statistical aspects of chaos: a review. In *Chaos and Networks: Statistical and Probabilistic Aspects: Proc. of a Workshop held at Sandbjerg Manor and the Department of Theoretical Statistics, Aarhus University, 25 April to 7 May 1992* (ed. O. E. Barndorff-Nielsen, D. R. Cox, J. L. Jensen & W. S. Kendall). London: Chapman & Hall.

Kendall, M. & Ord, J. K. 1990 *Time series.* London: Edward Arnold.

Silverman, B. W. 1986 *Density estimation for statistics and data analysis.* London: Chapman & Hall.

Skaug, H. J. & Tjøstheim, D. 1993*a* A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **80**, 591–602.

Skaug, H. J. & Tjøstheim, D. 1993*b* Nonparametric tests of serial independence. In *Developments in time series analysis.* The M. B. Priestley Birthday Volume (ed. T. Subba Rao). London: Academic Press.

Smith, L. A. 1988 Intrinsic limits in dimension calculations. *Phys. Lett.* A **133**, 283.

Stein, C. M. 1972 A bound for the error in the normal approximation to the distribution for a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 583–602. Berkeley: University of California Press.

Takens, F. 1981 On the numerical determination of the dimension of an attractor. In *Lecture Notes in Mathematics* **898**, (ed. D. A. Rand & L. S. Young), 366–381. New York: Springer Verlag.

Discussion

P. M. Robinson (*London School of Economics, U.K.*). First, in respect of Professor Subba Rao's reference to alternative methods based on higher-order spectra, let me note that these require moment conditions not required by estimates of the correlation integral which involve bounded functions. However, the methods based on higher-order spectra (to which Professor Subba Rao has made outstanding contributions) have a complementary role to play, in particular the spectral approach allows us to look at departures from independence at all lags, not just finitely many.

A difference between the central limit theory of Brock *et al.* and that of Dr Wolff is that they keep the bandwidth h fixed as sample size increases. As in most of the other recent work on testing for independence involving density estimates, he lets $h \to 0$. A difficulty with some of the recent methods (including one which I proposed and analysed) is that unless techniques such as sample splitting or trimming are introduced, there is a problem of degeneracy in the null limiting distribution. An advantage of the $h \to 0$ theory is that it can enable a relatively neat, and h-free, description of the consistent directions of the tests. Has Dr Wolff investigated conditions for consistency of his tests?

R. Wolff. I agree with Professor Robinson's response to Professor Subba Rao's comment. It would be interesting to determine if features of dependence which spectral analyses can detect have an equivalent revelation via the correlation integral, if at all, and, in respect of moment conditions, if the correlation integral's performance is superior.

Over the issue of consistency, I have not investigated this property. It may be that the so-called curse of dimensionality might confound it. Although that is purely a conjectural remark, it should be given due attention in the analysis.