Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 12128–12131, December 1996
Statistics

What is a linear process?

(chaos plus noise/ergodicity/Gaussian process/infinitely divisible law/nonlinear process)

PETER J. BICKEL AND PETER BÜHLMANN

Department of Statistics, University of California, Berkeley, CA 94720-3860

Contributed by Peter J. Bickel, August 1, 1996

ABSTRACT **We argue that given even an infinitely long data sequence, it is impossible (with any test statistic) to distinguish perfectly between linear and nonlinear processes (including slightly noisy chaotic processes). Our approach is to consider the set of moving-average (linear) processes and study its closure under a suitable metric. We give the precise characterization of this closure, which is unexpectedly large, containing nonergodic processes, which are Poisson sums of independent and identically distributed copies of a stationary process. Proofs of these results will appear elsewhere.**

1. Preliminary Description of Problems and Results

It has long been known, though perhaps not always appreciated, that it is impossible to test whether a set of observations comes from a "linear" ergodic or nonergodic Gaussian process since any nonergodic Gaussian process can be arbitrarily well approximated in a suitable metric by ergodic Gaussian processes, which are necessarily linear. We will present here a novel result that essentially any stationary process cannot be sharply distinguished from a linear process. Loosely, we consider the following problem: Given a partial realization $x_1, \ldots, x_n$ of a strictly stationary stochastic process $\{X_t\}_{t\in\mathbb{Z}}$, where $\mathbb{Z} = \{0, \pm1, \pm2, \ldots\}$, when can we conclude that the process is linear?

In recent years there has been a considerable interest in nonlinear time series analysis in the statistical, econometric, and engineering literatures (1–3). "Nonlinear" corresponds to many subclassifications, such as "bilinear" or "threshold autoregressive." Also, noisy chaotic processes defined by

$$X_t = f(X_{t-1}) + \varepsilon_t \ (t \in \mathbb{Z}),$$

where ε_t i.i.d. with $\mathbf{E}[\varepsilon_t] = 0$ and $f: \mathbb{R} \to \mathbb{R}$, define a subclass of nonlinear processes (for general f). But, at least linearity is fairly unambiguously specified. A linear stationary process $(X_t)_{t\in\mathbb{Z}}$ is usually described by

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \ (t \in \mathbb{Z}), \qquad \textbf{[1.1]}$$

where ε_t i.i.d. with $\mathbf{E}[\varepsilon_t] = 0$, $\mathbf{E}|\varepsilon_t|^2 < \infty$ and $\sum_{j=0}^{\infty}\psi_j^2 < \infty$. Such processes are also called moving-average (MA) processes. Here, we always assume existence of second moments. There is no loss of generality in assuming $\mathbf{E}[X_t] = 0$. Note that causal (minimum phase) autoregressive or ARMA processes are also representable as MA processes.

Given a finite stretch of a realization of a stationary process, one can try and test the hypothesis of linearity as stated in Eq. **1.1**. Such omnibus tests have been proposed, mainly by looking at higher order spectra (4, 5). But this hypothesis can be rejected only if alternatives are not well approximated by processes satisfying the hypothesis. The problem of testing H_0 about MA representation as in Eq. **1.1** leads then to the problem of studying the closure of the set of probability distributions of MA processes as given in Eq. **1.1** (MA closure).

The notion of a closed set requires the specification of a topology. We work here with the Mallows metric (6), also known as the Wasserstein metric, and with the stronger total variation metric. (For details, see sections 2.1 and 2.2.) We always identify real-valued stochastic processes, indexed by $\mathbb{Z}$, with their corresponding probability distributions; we then prefer to state our results in terms of stochastic processes.

We will argue in section 1.3 that the Mallows MA closure is exhausted by three types of processes. The first type is the set of stationary Gaussian processes with mean zero, i.e.,

$$S_1 = \{(X_t)_{t\in\mathbb{Z}};\ (X_t)_{t\in\mathbb{Z}} \text{ stationary Gaussian process with } \mathbf{E}[X_t] = 0\}.$$

The second type is the set of genuine MA processes, i.e.,

$$S_2 = \{(X_t)_{t\in\mathbb{Z}};\ X_t \text{ as defined in Eq. } \mathbf{1.1}\}.$$

The third type which arises is more surprising. We essentially can get Poisson sums of independent and identically distributed copies of stationary processes in the following sense. Denote by

$$(\xi_{t;1})_{t\in\mathbb{Z}},\ (\xi_{t;2})_{t\in\mathbb{Z}},\ \ldots,$$

a sequence of independent, real-valued, stationary processes with mean zero and finite second moments $\mathbf{E}|\xi_{t;1}|^2 = \sigma^2_{\xi;1}$, $\mathbf{E}|\xi_{t;2}|^2 = \sigma^2_{\xi;2}, \ldots$. Moreover, we construct for every $i \in \mathbb{N} = \{1, 2, \ldots\}$ a sequence of independent copies of $(\xi_{t;i})_{t\in\mathbb{Z}}$, namely

$$(\xi_{t;i,1})_{t\in\mathbb{Z}},\ (\xi_{t;i,2})_{t\in\mathbb{Z}},\ \ldots.$$

Thus we have constructed a sequence of processes

$$\{(\xi_{t;i,j})_{t\in\mathbb{Z}}\}_{i,j\in\mathbb{N}} \text{ independent processes over the index set}$$

$$i, j \in \mathbb{N}, (\xi_{t;i,1})_{t\in\mathbb{Z}}, (\xi_{t;i,2})_{t\in\mathbb{Z}}, \ldots \text{ i. i. d.}, \mathbf{E}[\xi_{t;i,j}] = 0, \mathbf{E}|\xi_{t;i,j}|^2 = \sigma^2_{\xi;i} < \infty. \qquad \textbf{[1.2]}$$

Let

$$N_1, N_2, \ldots \text{ independent, } N_i \sim \text{Poisson}(\lambda_i),\ \lambda_i \geq 0 \text{ for all } i \in \mathbb{N}. \qquad \textbf{[1.3]}$$

Then the third type is given by the following set of processes,

$$S_3 = \left\{ (X_t)_{t\in\mathbb{Z}};\ X_t = \sum_{i=1}^{\infty}\sum_{j=1}^{N_i} \xi_{t;i,j},\ (\xi_{t;i,j})_{t\in\mathbb{Z}},\ N_i \text{ satisfying } \mathbf{1.2}, \mathbf{1.3}, \text{ and} \sum_{i=1}^{\infty}\lambda_i\sigma^2_{\xi;i} < \infty \right\}.$$

Abbreviation: MA, moving average.

We make the convention that $\Sigma_{j=1}^{0} \xi_{t;i,j} = 0$. Elements of S_3 are typically nonergodic processes whose finite dimensional distributions are infinitely divisible non-Gaussian.

1.1. Nonergodic Limits and Separation Dilemma. In an informal way, the ergodic hypothesis postulates the equality of time-averages with averages over the elements in a probability space (in statistical mechanics, "phase-averages" in the phase space of a mechanical system). But the distinction between ergodic and nonergodic processes can be blurred.

Example 1.1: Consider the sequence of finite order MA processes,

$$X_t^{(n)} = \sum_{j=1}^{n} \xi_{j;1} U_{t-j,n} Z_{t-j,n} \ (t \in \mathbb{Z}),$$

with U_t i.i.d., $\mathbb{P}[U_t = 1] = 1 - \mathbb{P}[U_t = 0] = \lambda/n$ ($\lambda > 0$), Z_t i.i.d. $\sim t_5$, Student's t distribution with 5 degrees of freedom, and coefficients $(\xi_{j;1})_{j \in \mathbb{N}}$ which are a fixed realization of the Gaussian AR(1), $\xi_{j;1} = 0.9\xi_{j-1;1} + \eta_j$, η_j i.i.d. $\sim \mathcal{N}(0, 1)$.

For every $n \in \mathbb{N}$, these are ergodic MA processes of finite order n. But they exhibit a behavior which can be interpreted as nonergodic and "nonstationary," and which seems far from what one expects of a linear process. The reason is that they are close to a nonergodic member in S_3.

To illustrate the nonergodic phenomenon, we show in Fig. 1*A* nine realizations of sample size 500 of the process in Example 1.1 with $n = 50$. Fig. 1*A* tells in a quite impressive manner how different such realizations can be, and thus indicates that time-averages are not compatible with phase-averages over different realizations—i.e., nonergodic behavior. Fig. 1*B* shows one realization of sample size 5000 of the MA process in Example 1.1 with $n = 200$, now indicating nonstationarity. Different stretches of the sequence exhibit very different behaviors. This is the typical pattern for a time series with innovation outliers (7). Indeed, our model is an extreme case with innovations being either zero with probability $1 - \lambda/n$ or being a realization from a long-tailed distribution with probability λ/n. Note that outliers are with reference to the Gaussian distribution; it is the nonnormality of innovations which can lead to MA processes being close to a process in S_3.

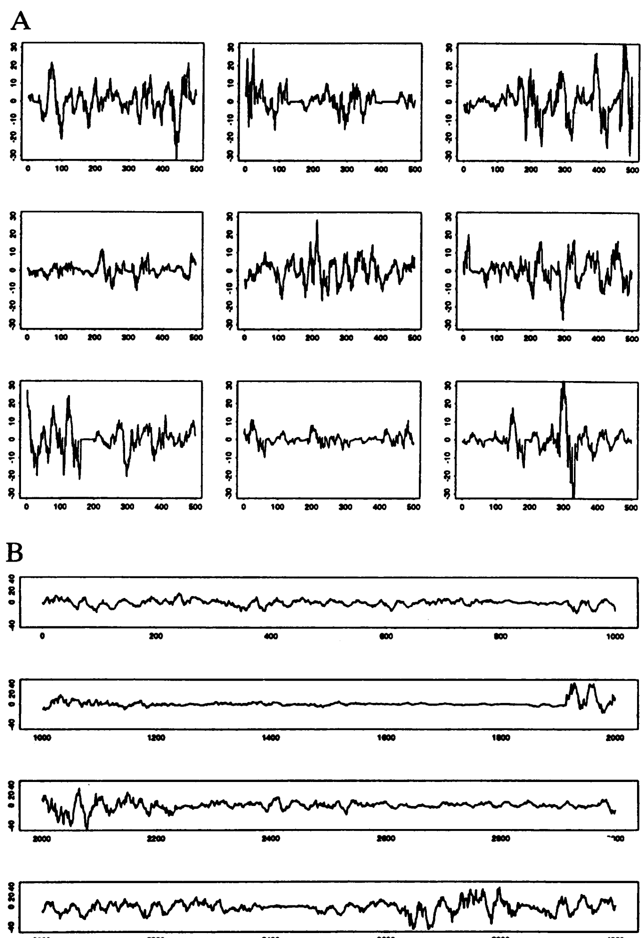

FIG. 1. (*A*) Nine realizations of Example 1.1 with $n = 50$, $\lambda = 3$. (*B*) One long realization of Example 1.1 with $n = 200$, $\lambda = 5$.

Example 1.1 is a special case of a very disturbing subclass of MA processes close to S_3. Given any, even infinitely long, realization $(\xi_t)_{t \in \mathbb{Z}}$ from any stationary process, consider the process $(X_t)_{t \in \mathbb{Z}} \in S_3$, where $X_t = \Sigma_{j=1}^{N} \xi_{t;j}$ ($t \in \mathbb{Z}$) with $N \sim$ Poisson(1) and $(\xi_{t;1})_{t \in \mathbb{Z}} = (\xi_t)_{t \in \mathbb{Z}}$, $(\xi_{t;2})_{t \in \mathbb{Z}}, \ldots$ independent identically distributed copies. It can be shown that this process is an element of the MA closure, compare also with Fact 1.4 in section 1.3. Since $\mathbb{P}[N = 1] = e^{-1} > 0.36$, we obtain $\mathbb{P}[X_t = \xi_t$ for all $t \in \mathbb{Z} | (\xi_t)_{t \in \mathbb{Z}}] > 0.36$. Summarizing, we have the following separation dilemma.

FACT 1.1. *Given any stationary process* $(\xi_t)_{t \in \mathbb{Z}}$, *there exists a nonergodic, stationary process* $(X_t)_{t \in \mathbb{Z}}$ *in the MA closure, which is an element of* S_3 *and has with positive probability exactly the same sample path as* $(\xi_t)_{t \in \mathbb{Z}}$. *More precisely,*

$$\mathbb{P}[X_t = \xi_t \textit{ for all } t \in \mathbb{Z} | (\xi_t)_{t \in \mathbb{Z}}] > 0.36 \textit{ almost surely.}$$

Details are given in Theorem 2.2. This separation dilemma is of the same nature as de Finetti's Theorem which can be thought of as stating the impossibility of distinguishing exchangeable from i.i.d. sequences (8, pp. 40–42).

In terms of the whole stochastic process, rather than a sample path, we have the following.

FACT 1.2. *The MA closure does not contain the set of ergodic, stationary processes.*

To show the validity of Fact 1.2, it is sufficient to give an example.

Example 1.2: Consider the stationary Markov chain $(X_t)_{t \in \mathbb{Z}}$, given by $X_t \in \{0, 1\}$ with $\mathbb{P}[X_1 = 0] = \mathbb{P}[X_1 = 1] = 1/2$, $\mathbb{P}[X_1 = 0 | X_0 = 0] = \mathbb{P}[X_1 = 0 | X_0 = 1] = \pi$, $0 < \pi < 1/2$. Then $(X_t)_{t \in \mathbb{Z}}$ is ergodic. Moreover, the probability distribution of X_t is not divisible, since the convolution of two nondegenerate distributions would place mass on at least three points, whereas X_t is only binary. Hence, the distribution of X_t cannot be approximated by any MA process and $(X_t)_{t \in \mathbb{Z}}$ can therefore not be an element of the MA closure.

It is possible to construct an ergodic, stationary process, with marginal distributions having a density with respect to Lebesgue measure, which is not an element of the MA closure (see ref. 15).

There are probably many ergodic, stationary processes, which are not elements of the MA closure. A possible candidate is the bilinear process, given by

$$X_t = -0.4X_{t-1} + 0.4X_{t-1}\varepsilon_{t-1} + \varepsilon_t (t \in \mathbb{Z}),$$

where ε_t i.i.d. $\sim \mathcal{N}(0, 1)$ (see figure 3.10 in ref. 9).

This process is stationary and ergodic (10). It is also immediate that the process is non-Gaussian. As argued in Subba Rao and Gabr (table 3.2 and figure 3.3 in ref. 9), this bilinear process is not representable as a moving average process. However, the MA closure also contains the class S_3 some of whose members may be ergodic.

1.2. The Testing Dilemma. There is considerable interest in testing the hypothesis that an observed time series is a linear process. Several authors propose different procedures for testing the hypothesis of MA representation (4, 5) and of autoregressive representation (11).

Consider the problem of distinguishing between the hypothesis $H_0 : (X_t)_{t\in\mathbb{Z}}$ is a linear process against the alternative $H_A : (X_t)_{t\in\mathbb{Z}}$ is a specific stationary process (not approximable by H_0 processes). Do there exist critical regions C_n for rejecting H_0, such that $\mathbb{P}_{H_0}[(X_1, \ldots, X_n) \in C_n] \to \alpha > 0$ and $\mathbb{P}_{H_A}[(X_1, \ldots, X_n) \notin C_n] \to 0$ as $n \to \infty$. That is, can one distinguish perfectly between H_0 and H_A at any level of significance α? Fact 1.1 can be restated as follows.

FACT 1.3. *In testing the hypothesis* H_0 *about MA representation against any fixed one-point alternative* H_A *about a nonlinear, stationary process, there is no test with asymptotic significance level* $\alpha < 0.36$ *having limiting power 1 as the sample size tends to infinity.*

1.3. Exhausting the MA Closure. The sets S_1, S_2, S_3 are not rich enough to exhaust the Mallows MA closure. To achieve this, we need sums of processes of the different types. We introduce an adding operation for processes and define

$(X_t)_{t\in\mathbb{Z}} \oplus (Y_t)_{t\in\mathbb{Z}}$ is the process $(X_t + Y_t)_{t\in\mathbb{Z}}$,

where the processes $(X_t)_{t\in\mathbb{Z}}$ and $(Y_t)_{t\in\mathbb{Z}}$ are independent.

We then set

$$S_i \oplus S_j = \{(X_t)_{t\in\mathbb{Z}} \oplus (Y_t)_{t\in\mathbb{Z}};\ (X_t)_{t\in\mathbb{Z}} \in S_i,\ (Y_t)_{t\in\mathbb{Z}} \in S_j\},\ i, j \in \{1, 2, 3\},$$

and make the common convention that all S_i ($i = 1, 2, 3$) also contain the null element $X_t \equiv 0$ for all $t \in \mathbb{Z}$.

The representation of a process as a $\oplus$-sum of elements in S_i ($i = 1, 2, 3$) is not unique even in the Gaussian case.

FACT 1.4. *The closure of the set of MA processes is given by*

$$\{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}.$$

Details are given in Theorem 2.1. Mallows (12) argues that a linear process such as in Eq. **1.1** is close to a Gaussian process if $\max_{j\geq 0} |\psi_j|$ is small. This is no longer true if one considers sequences $\{(X_{t,n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}$ of linear processes with coefficients $\psi_{j,n}$ as above and variables $\varepsilon_{t,n}$, which are i.i.d. but depend on n. Then, if $\max_{j\geq 0}|\psi_{j,n}| \to 0$ ($n \to \infty$) the process $(X_{t,n})_{t\in\mathbb{Z}}$ can have marginal distributions close to a non-Gaussian (not purely Gaussian) infinitely divisible law. Our result is in the spirit of Lévy (13) and uses his arguments. He showed that every continuous time process $(X_t)_{t\in\mathbb{Z}}$ with independent increments must have an infinitely divisible law and that such processes can be realized by a process with independent time homogeneous increments.

2. Precise Formulations

We consider real-valued, stationary processes $(X_t)_{t\in\mathbb{Z}}$ with expectation zero and finite variances. Thus, an appropriate probability space is $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}, \mathcal{P})$, where $\mathcal{B}$ denotes the Borel σ-field on $\mathbb{R}^{\mathbb{Z}}$ and $\mathcal{P}$ a class of stationary probability measures on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B})$, such that for every $P \in \mathcal{P}$,

$$\mathbf{E}_P[X] = \int_{\mathbb{R}} x d(P \circ \pi_0^{-1})(x) = 0,$$

$$\mathbf{E}_P|X|^2 = \int_{\mathbb{R}} x^2 d(P \circ \pi_0^{-1})(x) < \infty,$$

where $\pi_{t_1}, \ldots, {}_{t_m} : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^m$, $(x_t)_{t\in\mathbb{Z}} \mapsto (x_{t_1}, \ldots, x_{t_m})$, $t_1, \ldots, t_m \in \mathbb{Z}$.

We always identify a probability measure $P \in \mathcal{P}$ with its corresponding real-valued stochastic process.

It is possible to metrize the space $\mathcal{P}$ with a metric d (see sections 2.1 and 2.2). The closure with respect to the metric d of sets in $\mathcal{P}$, or equivalently of stationary real-valued stochastic processes with distributions in $\mathcal{P}$, is defined in the usual topological sense. We are particularly interested in the closure of MA processes (MA closure). Thus, we will consider sequences

$$\left\{\left(X_{t,n} = \sum_{j=0}^{\infty} \psi_{j,n}\varepsilon_{t-j,n}\right)_{t\in\mathbb{Z}}\right\}_{n\in\mathbb{N}}. \qquad [2.1]$$

2.1. Mallows Metric. We define the Mallows metric d_2 on $\mathcal{P}$, by

$$d_2(P_1, P_2) = \sum_{m=1}^{\infty} d_2^{(m)}(P_1 \circ \pi_{1,\ldots,m}^{-1}, P_2 \circ \pi_{1,\ldots,m}^{-1})2^{-m}, \quad P_1, P_2 \in \mathcal{P},$$

where $d_2^{(m)}(P_1 \circ \pi_{1,\ldots,m}^{-1}, P_2 \circ \pi_{1,\ldots,m}^{-1}) = \inf\{(\mathbf{E}\|X - Y\|^2)^{1/2}\}$ when the infimum is taken over all jointly distributed $(X, Y) \in \mathbb{R}^{2m}$ having marginals $P_1 \circ \pi_{1,\ldots,m}^{-1}$ and $P_2 \circ \pi_{1,\ldots,m}^{-1}$; $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^m$.

The following characterization is useful. Let $P_n, P \in \mathcal{P}$ and denote by $\Rightarrow$ weak convergence of probability measures. Then,

$$d_2(P_n, P) \to 0 \ (n \to \infty)$$

is equivalent to the following two statements

$P_n \circ \pi_{t_1,\ldots,t_m}^{-1} \Rightarrow P \circ \pi_{t_1,\ldots,t_m}^{-1}$ $(n \to \infty)$ for all $t_1, \ldots, t_m \in \mathbb{Z}, m \in \mathbb{N}$,

$$\int_{\mathbb{R}} x^2 d(P_n \circ \pi_0^{-1})(x) \to \int_{\mathbb{R}} x^2 d(P \circ \pi_0^{-1})(x) \ (n \to \infty),$$

that is, all finite dimensional distributions at $t_1, \ldots, t_m$ converge weakly and the variance of the marginal at any time point t converges (see ref. 14). We also use the notation for the corresponding processes, $d_2((X_{t,n})_{t\in\mathbb{Z}}, (X_t)_{t\in\mathbb{Z}}) = d_2(P_n, P)$, where $(X_{t;n})_{t\in\mathbb{Z}} \sim P_n$, $(X_t)_{t\in\mathbb{Z}} \sim P$.

2.2. Variation Metric. The question about distinguishing perfectly between two stationary processes requires a stronger metric than the Mallows d_2. The variation metric allows a precise formulation.

As before, let $P_1, P_2 \in \mathcal{P}$ and define the variation metric as

$$d_V(P_1, P_2) = \sum_{m=1}^{\infty} d_V^{(m)}(P_1 \circ \pi_{1,\ldots,m}^{-1}, P_2 \circ \pi_{1,\ldots,m}^{-1})2^{-m},$$

where $d_V^{(m)}(P_1 \circ \pi_{1,\ldots,m}^{-1}, P_2 \circ \pi_{1,\ldots,m}^{-1}) = \sup\{|P_1 \circ \pi_{1,\ldots,m}^{-1}[A] - P_2 \circ \pi_{1,\ldots,m}^{-1}[A]|; A \in \mathcal{B}(\mathbb{R}^m)\}$, $\mathcal{B}(\mathbb{R}^m)$ the Borel σ-algebra of $\mathbb{R}^m$. This definition reflects the nonuniform convergence of finite dimensional distributions in the variation metric. Here we do not require convergence of second moments. Distinguishing perfectly is characterized as follows. Let P_1, P_2 be ergodic probability measures in $\mathcal{P}$. Then

$d_V(P_1, P_2) > 0$ if and only if there exist test functions

$$\varphi_m : \mathbb{R}^m \to \mathbb{R}, 0 \leq \varphi_m \leq 1, \text{ such that } \mathbf{E}_{P_1}[\varphi_m(X_1, \ldots, X_m)] \to 0, \mathbf{E}_{P_2}[\varphi_m(X_1, \ldots, X_m)] \to 1 (m \to \infty).$$

2.3. Closure for MA Processes. We consider first the Mallows d_2 closure for MA processes, that is, sequences as defined in Eq. **2.1**. Without loss of generality we can scale the innovations and assume: (A): For every $n \in \mathbb{N}$, $(\varepsilon_{t,n})_{t \in \mathbb{Z}}$ is an i.i.d. sequence with

$$\mathbf{E}[\varepsilon_{t,n}] = 0, \ \mathbf{E}|\varepsilon_{t,n}|^2 = 1.$$

The following result describes the Mallows MA closure.

THEOREM 2.1. (i) *Consider a sequence of MA processes as defined in Eq.* **2.1** *converging in the* d_2 *sense, satisfying* (A) *and one of the following:*

(A1): $d_2^{(1)}(\varepsilon_{t,n}, \varepsilon_t) \to 0$ $(n \to \infty)$, *where* $(\varepsilon_t)_{t \in \mathbb{Z}}$ *is an i.i.d. sequence with* $\mathbf{E}[\varepsilon_t] = 0$.

(A2): $max_{j \geq 0}|\psi_{j,n}| \to 0$ $(n \to \infty)$.

Then, the d_2 *limit of such a sequence is in* $\{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}$.

(ii) *Every element of* $\{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}$ *can be obtained as a* d_2 *limit of a sequence of MA processes as defined in Eq.* **2.1,** *satisfying* (A) *and* (A1) *or* (A2).

Example 1.1 describes a sequence of MA processes with d_2 limit in S_3. This example can be modified so that the sequence of MA processes also converges in the variation metric to a d_V limit in S_3. This is needed in the following theorem, which has as a consequence that we can never distinguish perfectly between any stationary processes and MA processes even though there are such processes that cannot be approximated arbitrarily closely by MA processes.

THEOREM 2.2. *The MA closure with respect to the variation metric* d_V *has the following features.*

(i) *Let* $(\xi_t)_{t \in \mathbb{Z}}$ *be any stationary process such that for all* $m \in \mathbb{N}$, *the distributions of* $(\xi_1, \ldots, \xi_m)$ *have densities with respect to Lebesgue measure. Then, there exists a process* $(X_t)_{t \in \mathbb{Z}} \in S_3$, *which is an element of the MA closure with respect to the variation metric* d_V, *such that*

$$\mathbb{P}[X_t = \xi_t \textit{ for all } t \in \mathbb{Z} | (\xi_t)_{t \in \mathbb{Z}}] > 0.36 \textit{ almost surely.}$$

(ii) *There exist ergodic, stationary processes as in* (i) *which are not elements of the MA closure with respect to the variation metric* d_V.

The proofs of Theorem 2.1 and 2.2 are given in Bickel and Bühlmann (15). We have looked here at MA processes of infinite order. All our results are also true for sequences of finite (generally unbounded) order MA processes, which are more common in statistical modeling.

3. Discussion

The basic implication of our results is that any stationary process cannot be sharply distinguished from a high enough order MA process. Our proofs in Bickel and Bühlmann (15) show that a high order is a necessity to approximate an arbitrary ergodic, stationary process in the sense of Fact 1.1 and Theorem 2.2.

However, as can be noted from Fig. 1 the phenomenon is quite noticeable even for ratios of number of parameters to observations as low as 0.1. Note that purely chaotic processes do not fall under Theorem 2.2 since $(\xi_1, \ldots, \xi_m)$ do not have a density for m sufficiently large. However, by adding an arbitrarily small amount of white noise to any stationary process including purely chaotic ones we produce a process which can not be distinguished perfectly from an MA process of high order.

We thank David Freedman for helpful comments. P.J.B. was supported in part by National Security Agency Grant MDA 904-94-H-2020 and National Science Foundation Grant 95049555. P.B. was supported in part by the Swiss National Science Foundation.

1. Tong, H. (1990) *Non-Linear Time Series: A Dynamical System Approach* (Oxford Univ. Press, New York).
2. Priestley, M. B. (1988) *Non-Linear and Non-Stationary Time Series Analysis* (Academic, London).
3. Gershenfeld, N. A. & Weigend, A. S. (1994) *Time Ser. Prediction: Forecasting the Future and Understanding the Past* (Addison–Wesley, Reading, PA).
4. Subba Rao, T. & Gabr, M. M. (1980) *J. Time Ser. Anal.* **1,** 145–158.
5. Hinich, M. (1982) *J. Time Ser. Anal.* **3,** 169–176.
6. Mallows, C. L. (1972) *Ann. Math. Stat.* **43,** 508–515.
7. Kleiner, B., Martin, R. D. & Thomson, D. J. (1979) *J. R. Stat. Soc. B* **41,** 313–351.
8. Hartigan, J. A. (1983) *Bayes Theory* (Springer, New York).
9. Subba Rao, T. & Gabr, M. M. (1984) *An Introduction to Bispectral Analysis and Bilinear Time Series Models,* Lecture Notes in Statistics (Springer, Heidelberg), Vol. 24.
10. Akamanam, S. I., Rao, M. Bhaskara & Subramanyam, K. (1986) *J. Time Ser. Anal.* **7,** 157–163.
11. Hjellvik, V. & Tjøstheim, D. (1995) *Biometrika* **82,** 351–368.
12. Mallows, C. L. (1967) *J. Appl. Prob.* **4,** 313–329.
13. Lévy, P. (1948) *Processus Stochastiques et Mouvement Brownien* (Gauthier-Villars, Paris).
14. Bickel, P. J. & Freedman, D. A. (1981) *Ann. Stat.* **9,** 1196–1217.
15. Bickel, P. J. & Bühlmann, P. (1997) *J. Theor. Probab.*, in press.