The Annals of Statistics
1982, Vol. 10, No. 3, 786–794

TIME SERIES DISCRIMINATION BY HIGHER ORDER CROSSINGS

BY BENJAMIN KEDEM AND ERIC SLUD[1]

University of Maryland

A new methodology is proposed for discrimination among stationary time-series. The time series are transformed into binary arrays by clipping (retaining only the signs of) the jth difference series, $j = 0, 1, 2, \cdots$. The degeneracy of clipped jth differences is studied as j becomes large. A new goodness of fit statistic is defined as a quadratic form in the counts of axis-crossings by each of the first k differences of the series. Simulations and the degeneracy of high-order differences justify fixing k no larger than 10 for many processes. Empirical simulated distributions (with $k = 9$) of the goodness of fit statistic suggest a gamma approximation for its tail probabilities. Illustrations are given of discrimination between simple models with the new statistic.

1. Introduction. In many areas of application of time series methods, data is generated in repeated cycles by a machine or organism whose operation is monitored for change to a "malfunction" mode. Examples include electrocardiographic and EEG series and engineering tests of devices and mechanical structures under repeated loading. Common to these diverse areas is the need to develop a broad criterion of goodness of fit for time series (stochastic processes) applicable not only to strictly and wide-sense stationary processes but to the nearly-periodic random waveforms arising in engineering contexts.

Nearly all current tests of fit for time series relate to the parametric classes of Auto-Regressive Moving Average (ARMA) or Gaussian processes, in which, respectively, sums of squared residuals or residual-correlations (Box and Jenkins, 1970) and likelihood-ratio statistics (Grenander, 1950; Whittle, 1951) are used. A method of large-sample discrimination between non-nested Gaussian ARMA models (and between AR and sinusoid-plus-white noise models) has been worked out by Walker (1967), using the idea of Cox's (1962) "tests of separate families." Successful discrimination among "locally stationary" Gaussian models for electroencephalograph (EEG) and machine-vibration data has been reported by W. Gersch and co-workers (1979a, b).

Ad hoc and problem-based methods of discriminating time series almost necessarily do without parametric assumptions. For example, John, *et al* (1980) discriminate EEG data which is patently nonstationary by means of the power contained in specified frequency bands. In the area of automatic speech recognition, there are widespread methods of identification (reviewed by Niederjohn and Castelez, 1978) based on the number and durations of times when the speech-waveform is above a fixed level u. Other applications of level-crossings analysis are given by Becker (1978) and Schiess (1979). Cramér and Leadbetter (1963) give theoretical results on level crossings for stationary Gaussian processes, and Hinich (1967), Brillinger (1968), and Kedem (1979) discuss consistent parameter estimation from stationary Gaussian data clipped at the mean (i.e., data recorded only as above or below the known mean).

Each of the methods of discrimination mentioned above employs sample time-averages, therefore implicitly requiring some assumption of stationarity. All the analysis of the present paper will also presuppose (strict or wide-sense) stationary second order random sequences, but since we wish to avoid parametric assumptions, we restrict ourselves to discriminating via counts of features in the time domain.

Received January 1980; revised January 1982.

[1] The research of both authors was supported by the Air Force Directorate of Mathematical and Information Sciences under contract AFOSR F49620-79-C-0095.

AMS 1980 *subject classifications.* Primary 62M10; secondary 62M07, 62H30.

Key words and phrases. Stationary time series, goodness of fit, discrimination, higher-order crossings, level crossings, clipped signal.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

Our procedures concern level crossings of (higher-order) differences of a time series, and are most naturally expressed through the following representation of random sequences, which also serves to establish our notation.

Let $\mathbf{Z} = \{Z_t\}_{t=-\infty}^{\infty}$ be a (stationary) discrete-time process. We denote by ∇ the backward-difference operator acting on the entire sequence $\mathbf{Z}$, yielding tth coordinate $(\nabla\mathbf{Z})_t = Z_t - Z_{t-1}$; similarly $(\nabla^2\mathbf{Z})_t = (\nabla(\nabla\mathbf{Z}))_t = Z_t - 2Z_{t-1} + Z_{t-2}$, etc. The clipping operator U is defined by

$$(U\mathbf{Z})_t = \begin{cases} 1 & \text{if } Z_t \geq 0, \\ 0 & \text{if } Z_t < 0. \end{cases}$$

Then $\mathbf{X}_t = \mathbf{X}_t(\mathbf{Z}) = (X_t^{(1)}, X_t^{(2)}, \cdots)'$ for $t = 0, \pm 1, \pm 2, \cdots$ is the series of binary vectors in $\{0, 1\}^{\mathbb{Z}}$, *the binary differential representation* of $\mathbf{Z}$, defined for $j \geq 1$ by

$$X_t^{(j)} \equiv (U\nabla^{j-1}\mathbf{Z})_t \equiv \begin{cases} 1 & \text{if } (\nabla^{j-1}\mathbf{Z})_t \geq 0, \\ 0 & \text{if } (\nabla^{j-1}\mathbf{Z})_t < 0. \end{cases}$$

We further define, for $k \geq 1$ and large N, the truncated array

$$\mathbf{X}_N^{(k)} \equiv \mathbf{X}_N^{(k)}(\{Z_t\}) = \{(X_t^{(1)}, \cdots, X_t^{(k)})' : 1 \leq t \leq N\}$$

This is the form in which we encode and discriminate among series $\{Z_t\}$ and on which our goodness of fit statistic will be based. The special statistics $D_{j,N} \equiv \sum_{t=1}^{N-1} [X_{t+1}^{(j)} \neq X_t^{(j)}]$, the number of axis-crossings of $\nabla^{j-1}Z_t$ as t runs from 1 to N, will be called the *higher-order crossings* (of order $j \geq 1$) of the series $\{Z_t\}$.

There are applications of time series (especially in the social sciences) where the qualitative information contained in a truncated binary-differential representation is the most one should hope to forecast. In such instances, the truncated binary-differential representation corresponding to a fitted ARIMA model can be considered as a model for $\mathbf{X}_N^{(k)}$ itself, giving a less detailed but more robust description. Prediction and modeling of $\mathbf{X}_N^{(k)}$ remain topics for further research. The present paper is concerned only with discrimination via higher-order crossings.

In Section 2 below, we show that the degeneracy resulting from repeated differencing of wide-sense stationary sequences has important special consequences for strictly stationary second-order sequences. Then Section 3 justifies and illustrates the use of a particular goodness of fit test statistic ψ^2 constructed from counts of higher-order crossings. For further details on implementing one- and two-sample tests of fit with ψ^2, and for comparisons under ARMA models with other tests of fit, see Kedem and Slud (1981).

The apparent restriction of our methods to stationary time series is partially removed in Slud (1981), which extends all the results of the present work to a class of non-stationary process which includes the (nearly) periodic waveforms with stationary noise. Other extensions of higher-order crossings methods, to the discrimination of planar textures, have been explored in a report of Kedem (1981).

2. The higher-order crossings theorem. Suppose that $\mathbf{Z} = \{Z_t\}_{t=-\infty}^{\infty}$ (t integer-valued) is a zero-mean, real-valued wide-sense stationary process. Our general reference for this paragraph is Doob (1953). It is well known that the covariance function $r(k)$ of $\mathbf{Z}$ corresponds to a uniquely defined spectral distribution function F on $(-\pi, \pi]$ so that

$$E(Z_0 Z_k) \equiv r(k) \equiv \int_{-\pi}^{\pi} e^{ik\lambda}\, dF(\lambda).$$

This implies the existence of a complex-valued process $\xi(\lambda)$ with orthogonal increments on $(-\pi, \pi]$ such that $Z_t = \int_{-\pi}^{\pi} e^{i+\lambda}\, d\xi(\lambda)$ in the mean-square sense, where $dF(\lambda) = E|d\xi(\lambda)|^2$.

Again, for $j \geq 1$, we define the series $\mathbf{X}^{(j)} \equiv \{X_t^{(j)}\}$ by $\mathbf{X}^{(j)} \equiv U\nabla^{j-1}\mathbf{Z}$; but now define $\mathbf{Y}^{(j)} = \{Y_t^{(j)}\} \equiv \nabla^{j-1}\mathbf{Z}$ and $\bar{\mathbf{Y}}^{(j)} = \{\bar{Y}_t^{(j)}\} \equiv c_j^{-1}\mathbf{Y}^{(j)}$, where $c_j > 0$ is chosen so that $\operatorname{Var} \bar{Y}_t^{(j)} = 1$. We normalize $\mathbf{Y}^{(j)}$ because the clipped sequence $\mathbf{X}^{(j)}$ does not depend on the amplitudes of $\nabla^{j-1}\mathbf{Z}$; hence $\mathbf{X}^{(j)} = U\bar{\mathbf{Y}}^{(j)}$. Since ∇^j is a linear filter, we can express

$$Y_t^{(j+1)} = \int_{-\pi}^{\pi} (1 - e^{-i\lambda})^j e^{i\lambda t}\, d\xi(\lambda)$$

and

$$c_{j+1}^2 = \operatorname{Var}(Y_t^{(j+1)}) = \int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{2j}\, dF(\lambda).$$

We further define $F_j(\cdot)$ to be the spectral distribution function of $\{\bar{Y}_t^{(j+1)}\}$, corresponding to the symmetric measure ν_j on $(-\pi, \pi]$. Thus

$$dF_j(\lambda) = \nu_j(d\lambda) = |1 - e^{-i\lambda}|^{2j}\, dF(\lambda) \Big/ \int_{-\pi}^{\pi} |1 - e^{-i\tau}|^{2j}\, dF(\tau).$$

In the frequency domain, the operators ∇^j are high-pass filters pushing spectral mass toward the higher frequencies. That repeated application of such filters leads to oscillatory degeneracy appears to be well known (cf. Grenander and Rosenblatt, 1957, Section 3.3; Anderson, 1971, Section 7.5.5). In this section, we obtain more precise information on this degeneracy for strictly stationary processes. The following proposition summarizes the asymptotics for large j of the spectral effect of ∇^j operating on wide-sense stationary $\mathbf{Z}$.

PROPOSITION 2.1. *Let $\alpha \in [0, \pi]$ be* max(support(F)) = *largest point of increase of F in $[0, \pi]$. Then as $j \to \infty$, $\nu_j \to_w \frac{1}{2}\delta_{-\alpha} + \frac{1}{2}\delta_\alpha$ if $\alpha < \pi$, and $\nu_j \to_w \delta_\pi$ if $\alpha = \pi$, where δ_u denotes point mass at u.*

PROOF OF PROPOSITION. By assumption, F has point of increase only in $[-\alpha, \alpha]$. Moreover $|1 - e^{-i\lambda}|^2 = 2(1 - \cos\lambda)$ is strictly increasing in $[0, \pi]$. Also recall that ν_j is a symmetric measure on $(-\pi, \pi]$.

Now for $\lambda \in [-\alpha + \varepsilon, \alpha - \varepsilon]$, where $\varepsilon > 0$, we have by monotonicity

$$|1 - e^{-i\lambda}|^2 \le \rho(\varepsilon) \equiv 2\{1 - \cos(\alpha - \varepsilon)\} < 2(1 - \cos\alpha).$$

Obviously $|1 - e^{-i\lambda}|^2 \ge \rho(\varepsilon)$ for $\lambda \in [-\alpha, -\alpha + \varepsilon) \cup (\alpha - \varepsilon, \alpha]$. But $dF(\lambda)$ is 0 off $[-\alpha, \alpha]$, hence

$$\nu_j([-\alpha + \varepsilon, \alpha - \varepsilon]) = \int_{-\alpha+\varepsilon}^{\alpha-\varepsilon} |1 - e^{-i\lambda}|^{2j}\, dF(\lambda) \Big/ \int_{-\alpha}^{\alpha} |1 - e^{-i\lambda}|^{2j}\, dF(\lambda).$$

Also for all $\varepsilon \in (0, \alpha)$

$$\int_{-\alpha}^{(-\alpha+\varepsilon)-} dF(\lambda) + \int_{(\alpha-\varepsilon)+}^{\alpha} dF(\lambda) \equiv \sigma(\varepsilon) > 0.$$

Therefore

$$\nu_j([-\alpha + \varepsilon, \alpha - \varepsilon]) \le \rho^j(\varepsilon)\{r(0) - \sigma(\varepsilon)\} \Big/ \left\{ \int_{-\alpha}^{[-\alpha+(\varepsilon/2)]-} + \int_{[\alpha-(\varepsilon/2)]+}^{\alpha} |1 - e^{-i\lambda}|^{2j}\, dF(\lambda) \right\}$$

$$\le \frac{\rho^j(\varepsilon)\{r(0) - \sigma(\varepsilon)\}}{\rho^j\left(\dfrac{\varepsilon}{2}\right)\sigma\left(\dfrac{\varepsilon}{2}\right)} \to 0$$

exponentially as $j \to \infty$, for all $\varepsilon > 0$, because $\rho(\varepsilon/2) > \rho(\varepsilon)$. It follows that

$$\nu_j([-\alpha, -\alpha + \varepsilon) \cup (\alpha - \varepsilon, \alpha]) \to 1, \qquad j \to \infty.$$

Since ν_j is symmetric on $(-\pi, \pi)$, we conclude as $j \to \infty$, $\nu_j \to_w \frac{1}{2}\delta_{-\alpha} + \frac{1}{2}\delta_\alpha$ if $\alpha < \pi$, $\nu_j \to_w \delta_\alpha$ if $\alpha = \pi$. Moreover, it follows from the definition of weak convergence that

$$\operatorname{Cov}(\bar{Y}_t^{(j)}, \bar{Y}_{t+m}^{(j)}) = \int_{-\pi}^{\pi} e^{im\lambda}\nu_j(d\lambda) \to \frac{1}{2}(e^{im\alpha} + e^{-im\alpha}) = \cos(m\alpha) \quad \text{as} \quad j \to \infty.$$

In the special case $\alpha = \pi$ we have

$$\operatorname{Cov}(\bar{Y}_t^{(j)}, \bar{Y}_{t+m}^{(j)}) \to (-1)^m, \qquad j \to \infty. \tag{2.1}$$ □

From (2.1) follows, uniformly for all integers t,

$$E\,|\,\bar{Y}_t^{(j)} - (-1)^m \bar{Y}_{t+m}^{(j)}|^2 \to 2 - 2(-1)^{2m} = 0, \qquad j \to \infty. \tag{2.2}$$

At this point we encounter a technical problem which must be disposed of. We are interested in the behavior of $\{X_t^{(j)}\}$ for large values of j and strictly stationary $\mathbf{Z}$. In order to avoid difficulties associated with positive probability for the event $[Y_t^{(j)} = 0]$, we appeal to the following lemma, which may be of independent interest.

LEMMA 2.2. *If* $\mathbf{Z} = \{Z_t\}_{t=-\infty}^{\infty}$ *is a strictly stationary sequence such that*

$$P(Z_t \equiv Z_0 \; \forall t) = 0, \quad \textit{then} \quad \limsup_{n\to\infty} P(\{\omega : (\nabla^n \mathbf{Z})_0(\omega) = 0\}) = 0.$$

PROOF. We suppose for some sequence n_j tending to ∞ that for $j \geq 1$, $P((\nabla^{n_j}\mathbf{Z})_0 = 0) \geq \delta > 0$. By Fatou's Lemma $P((\nabla^{n_j}\mathbf{Z})_0 = 0$ for infinitely many $j) \geq \delta$. By Birkhoff's Ergodic Theorem, as $N \to \infty$,

$$N^{-1}\operatorname{card}\{1 \leq k \leq N : (\nabla^{n_j}\mathbf{Z})_k = 0 \quad \text{for infinitely many} \quad j \geq 1\}$$

converges with probability 1 to a random variable A with expectation $\geq\delta$. On the set $\{\omega : A(\omega) > 0\}$, for sufficiently large i the sequence $\{(\nabla^{n_i}\mathbf{Z})_t(\omega)\}_{t=1}^{\infty}$ is *overdetermined* by the value $(\nabla^{n_i}\mathbf{Z})_0(\omega)$ and the relations $(\nabla^{n_j}\mathbf{Z})_k = 0$ which occur for $j \geq i$, $t \geq 1$. Hence $A(\omega) > 0$ implies $\nabla^m Z_t(\omega) \equiv 0$ for some $m \geq 1$ and all $t \geq 1$. By assumption, the latter event has probability 0, since stationarity implies a.s. that for positive t, Z_t cannot be polynomial of degree at least 1. Therefore there can be no $\{n_j\}$ and $\delta > 0$ as above, and the lemma is proved. □

Whenever $\mathbf{Z}$ is wide-sense stationary with mean 0, has spectral df F for which π is a point of increase, and satisfies $\limsup_{n\to\infty} P(\nabla^n \mathbf{Z}_0 = 0) = 0$, (2.2) implies that

$$P[X_{t+m}^{(j)} \neq (-1)^m X_t^{(j)} + \tfrac{1}{2}\{1 - (-1)^m\}] \to 0 \quad \text{as} \quad j \to \infty,$$

uniformly in t. Similarly we have for each fixed $m \geq 1$,

$$\lim_{j\to\infty} P[X_{t+\ell}^{(j)} = (-1)^{\ell} X_t^{(j)} + \tfrac{1}{2}\{1 - (-1)^{\ell}\} \quad \text{for} \quad \ell = 1, \cdots, m] = 1. \tag{2.3}$$

We retain the same assumptions on $\mathbf{Z}$ throughout the next paragraph.

Recalling the notation $D_{j,N} \equiv \sum_{t=1}^{N-1} I[X_{t+1}^{(j)} \neq X_t^{(j)}]$, and observing that $X_t^{(j)} \neq X_{t+1}^{(j)}$ implies $X_{t+1}^{(j+1)} = X_{t+1}^{(j)}$ for $j \geq 1$, t any integer, we remark that surely

$$0 \leq D_{j,N} \leq N - 1 \quad \text{and} \quad D_{j+1,N} \geq D_{j,N} - 1. \tag{2.4}$$

It then follows from wide-sense stationarity for $\{Z_t\}$ by (2.3) with $m = 1$ that $E(D_{j,N}/N - 1) \to 1$ uniformly in N as $j \to \infty$. If we define $C_n = \limsup_{N\to\infty} D_{n,N}/N$ for $n \geq 1$, and make the crucial observation by (2.4) that $C_{n+1} \geq C_n$ a.s. while $0 \leq C_n \leq 1$, then we can conclude with probability 1, $\lim_{n\to\infty} C_n \equiv C \leq 1$ exists. By the Bounded Convergence Theorem $E(C) = \lim_{n\to\infty} E(C_n)$. By Fatou's Lemma $E(C_n) \geq \limsup_{N\to\infty} E(D_{n,N}/N)$, and the uniform convergence of $E(D_{n,N}/N - 1)$ to 1 as $n \to \infty$ implies $\limsup_{N\to\infty} E(D_{n,N}/N) \to 1$ as $n \to \infty$, and $E(C) = 1$. Therefore $C = 1$ a.s. We have proved:

THEOREM 2.3. *If* $\{Z_t\}_{t=-\infty}^{\infty}$ *is wide-sense stationary with mean* 0; *if* π *is a point of increase for its spectral distribution function* F; *and if also*

$$\limsup_{n\to\infty} P\{(\nabla^n \mathbf{Z})_t = 0\} = 0 \quad \textit{for all integers } t;$$

then

(a) $\mathrm{Corr}(X_t^{(n)}, X_{t+k}^{(n)}) \to (-1)^k$ *as* $n \to \infty$

(b) $\bar{D}_j \equiv \limsup_{N\to\infty} N^{-1} \sum_{t=1}^N I[X_{t+1}^{(j)} \neq X_t^{(j)}]$ *is almost surely increasing to* 1 *as* $j \to \infty$.

Of course, when $\{Z_t\}$ is strictly stationary, Lemma 2.2 can replace the assumption on $P\{(\nabla^n \mathbf{Z})_t = 0\}$, and Birkhoff's Ergodic Theorem implies $\lim_{N\to\infty} D_{j,N}/N = \bar{D}_j$ exists almost surely. If $\mathbf{Z}$ and hence $\mathbf{X}^{(j)}$ are ergodic, then $\bar{D}_j = P(X_{t+1}^{(j)} \neq X_t^{(j)})$ a.s.

HIGHER-ORDER CROSSINGS THEOREM. *Let* $Z = \{Z_t\}_{t=-\infty}^{\infty}$ *be strictly stationary, with finite variance, and* $P(Z_t \equiv Z_0\ \forall t) = 0$. *Suppose that* π *is a point of increase for the spectral distribution function* F *on* $(-\pi, \pi]$. *Then*

(i) *the* $0-1$ *valued processes* $\mathbf{X}^{(n)} \equiv \{X_t^{(n)}\}_{t=-\infty}^{\infty}$ *converge in distribution as* $n \to \infty$ *to the process* $\mathbf{a} = \{a_t\}_{t=-\infty}^{\infty}$ *defined by*

$$a_0 = \begin{cases} 1 & \textit{with prob. } \tfrac12 \\ 0 & \textit{with prob. } \tfrac12 \end{cases}$$

and $P(a_{k+1} = 1 - a_k$ *for all* $k = 0, \pm 1, \cdots) = 1$;

(ii) $$\{X_k^{(m)} : m = n, n+1, n+2, \cdots\} \to_{\mathscr{D}} \begin{cases} \mathbf{1} & \textit{with prob. } \tfrac12 \\ \mathbf{0} & \textit{with prob. } \tfrac12 \end{cases}$$

as $n \to \infty$ *for each integer* k, *where* $\mathbf{1} = (1, 1, \cdots)$ *and* $\mathbf{0} = (0, 0, \cdots)$;

(iii) $\lim_{j\to\infty} \lim_{N\to\infty} N^{-1} D_{j,N} = 1$ *with probability* 1.

PROOF. By Lemma 2.2 and Theorem 2.3, part (iii) and (2.3) have already been proved for the case $EZ_0 = 0$, and are not changed in replacing Z_t by $Z_t - EZ_0$. To prove weak-convergence statements, we recall that the natural product topology on $\Omega \equiv \{0, 1\}^{\mathbf{Z}}$ is generated from neighborhoods $V_M(\mathbf{y}) \equiv \{\mathbf{x} \in \Omega : x_k = y_k \text{ if } |k| \le M\}$, for $\mathbf{y} \in \Omega$. We define two special points $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ in Ω by

$$\alpha_t = \begin{cases} 1 & \text{if } t \text{ is odd,} \\ 0 & \text{if } t \text{ is even,} \end{cases} \qquad \alpha_t' = \begin{cases} 1 & \text{if } t \text{ is even} \\ 0 & \text{if } t \text{ is odd.} \end{cases}$$

Denoting by Q_n the law of $\{X_t^{(n)}\}_{t=-\infty}^{\infty}$ on $\{0, 1\}^{\mathbf{Z}}$, we translate (2.3) to mean for all M,

$$Q_n(V_M(\boldsymbol{\alpha}) \cup V_M(\boldsymbol{\alpha}')) = P\{\mathbf{X}^{(n)} \in V_M(\boldsymbol{\alpha}) \cup V_M(\boldsymbol{\alpha}')\} \to 1, \qquad n \to \infty.$$

Therefore any weak limit measure for $\{Q_n\}_{n=1}^{\infty}$ must be supported on $\{\boldsymbol{\alpha}, \boldsymbol{\alpha}'\}$ in Ω. That is, each weakly convergent subsequence Q_{n_k} has limit of the form $R_\beta = \beta \cdot \delta_\alpha + (1-\beta) \cdot \delta_{\alpha'}$, for some $\beta \in [0, 1]$.

Since $\mathbf{X}^{(n)}$ is stationary, the laws Q_n are all invariant under the left sequence-shift. Therefore all the R_β which appear as limits of Q_{n_k} are also stationary. But the sequence-shift carries $\boldsymbol{\alpha}$ into $\boldsymbol{\alpha}'$ and $\boldsymbol{\alpha}'$ into $\boldsymbol{\alpha}$, implying that the only shift-invariant R_β is $R_{1/2}$. Since $\{0, 1\}^{\mathbf{Z}}$ with the product topology is compact, every subsequence of $\{Q_n\}_{n=1}^{\infty}$ contains a weakly convergent subsequence which must converge to $R_{1/2} = (1/2)\delta_\alpha + (1/2)\delta_{\alpha'}$. Hence $Q_n \to_w R_{1/2}$ as $n \to \infty$, proving part (i).

Finally, to prove (ii) we recall that $X_{t+1}^{(n)} \neq X_t^{(n)}$ implies $X_{t+1}^{(n+1)} = X_{t+1}^{(n)}$. Then for fixed even m, $(X_t^{(n)}, X_{t+1}^{(n)}, \cdots, X_{t+m}^{(n)}) = (1, 0, 1, 0, 1, \cdots)$ implies $(X_{t+m}^{(n)}, X_{t+m}^{(n+1)}, \cdots, X_{t+m}^{(n+m)}) = (1, 1, \cdots, 1)$ and $(X_t^{(n)}, \cdots, X_{t+m}^{(n)}) = (0, 1, 0, 1, \cdots, 0)$ implies $X_{t+m}^{(n)}, \cdots, X_{t+m}^{(n+m)}) = (0, 0, \cdots, 0)$. Therefore part (ii) follows from (2.3) and part (i), and the proof is complete. □

3. The goodness of fit statistic. Throughout this section we assume the random sequence $\{Z_t\}$ is strictly stationary and satisfies the hypotheses of the Higher Order Crossings Theorem. In addition, for asymptotic normality of our higher-order crossings counts, we require that $\{Z_t\}$ be ϕ-mixing as in the hypothesis of Billingsley's (1968) Theorem 20.1.

By the theorems of Section 2, essentially all the predictive or discriminatory power in

$\{X_t^{(j)}: 1 \le t \le N, j \ge 1\}$ for large N is already contained in $\mathbf{X}_N^{(k)}$ for moderately large fixed k. Simulation results based on work with both linear and nonlinear process models corroborate our theorem and suggest that $\bar{D}_j = \lim_{N\to\infty} D_{j,N}/N$ is already above .9 when j is as large as 10 for a wide variety of examples. For this reason, we restrict attention in our statistical applications to $\mathbf{X}_N^{(k)}$ for $k \le 10$.

We recall from Section 2, formula (2.4), that $D_{j+1,N} \ge D_{j,N} - 1$ surely. Now fixing k once and for all, we define $\Delta_{j,N}$ for $1 \le j \le k$ by

$$\Delta_{j,N} = \begin{cases} D_{1,N} & \text{if } j = 1 \\ D_{j,N} - D_{j-1,N} & \text{if } j = 2, \cdots, k-1 \\ N - 1 - D_{k-1,N} & \text{if } j = k. \end{cases}$$

These $\Delta_{j,N}$ are asymptotically non-negative random variables, since

$$\Delta_{j,N}/N \to \begin{cases} P(X_{t+1}^{(1)} \ne X_t^{(1)}) & \text{if } j = 1 \\ P(X_{t+1}^{(j)} \ne X_t^{(j)}) - P(X_{t+1}^{(j-1)} \ne X_t^{(j-1)}) & \text{if } 2 \le j \le k-1 \\ P(X_{t+1}^{(k-1)} = X_t^{(k-1)}) & \text{if } j = k. \end{cases}$$

Let $m_{j,N} \equiv E\Delta_{j,N}$. By non-constancy and ergodicity of $\{Z_t\}$, $m_{1,N} > 0$ if $EZ_t = 0$. For $j \ge 2$, $m_{j,N} = 0$ would mean that the series $\{(\nabla^{j-2}\mathbf{Z})_t\}_t$ could change sign at t iff it had a local extremum at $t-1$, i.e., $X_t^{(j)} \equiv X_t^{(j-1)}$ for all t. This behavior is not excluded in principle even for $j = 2$, but is so distinctive as to render unnecessary a goodness of fit test between a model with $m_{j,N}$ approximately 0 for large N and data with $\Delta_{j,N} \gg 0$. From now on, we assume $m_{j,N} > 0$ for $j = 1, \cdots, k$.

For each $(\alpha_1, \cdots, \alpha_k)$, Billingsley's (1968, page 174) ϕ-mixing Central Limit Theorem with $\xi_t \equiv \sum_{j=1}^k \alpha_j I[X_{t+1}^{(j)} \ne X_t^{(j)}]$ implies $N^{-1/2} \sum_{j=1}^k \alpha_j(D_{j,N} - ED_{j,N})$ converges in distribution as $N \to \infty$ to a (possibly degenerate) normal variable. Hence $N^{-1/2}(\Delta_{1,N} - m_{1,N}, \cdots, \Delta_{k,N} - m_{k,N})$ is asymptotically multivariate-normal. We can now analogize $\Delta_{j,N}$ with the frequencies in a multinomial experiment to construct the goodness of fit statistic $\psi_N^2 \equiv \sum_{j=1}^k (\Delta_{j,N} - m_{j,N})^2/m_{j,N}$. The $m_{j,N}$ will be provided from extensive data (or model assumptions) about a null-hypothetical signature, and the $\Delta_{j,N}$ will typically derive from a newly collected series. When $m_{j,N}$ and $\Delta_{j,N}$ arise from the same $\mathscr{L}(\mathbf{Z})$ satisfying the assumptions of this section, the quadratic form ψ_N^2 is asymptotically for large N distributed as $\sum_{j=1}^{k-1} \lambda_j \eta_j^2$ where the η_j are independent standard normal variables, and the non-negative constants λ_j depend on $\mathscr{L}(\mathbf{Z})$. Of course many other goodness of fit statistics can be constructed as quadratic forms in $D_{j,N} - ED_{j,N}$, but ψ_N^2 is simple and avoids variance-covariance computations.

The empirical distribution of ψ_N^2 has been simulated using the IMSL Box-Mueller Gaussian random-number generator with $N = 1000$, $k = 9$ for many underlying processes $\{Z_t\}$—mainly ARMA processes, but also functions of ARMA processes and some unusual processes like those of Table 3.2, which follows. The shape of histograms for ψ^2 is always vaguely like a noncentral χ^2, with average values of ψ_{1000}^2 ($k = 9$) in the range 10–15 and sample variance from roughly 40 to 54. For example, for the sixth, tenth, and twelfth processes of Table 3.1 below, the respective simulated sample mean and variance pairs for ψ_{1000}^2 were (13.09, 52.78), (13.16, 43.25), and (13.78, 44.29). The object of our simulations has been to obtain reliable critical values of ψ^2, and indeed almost all our simulations with many processes $\mathbf{Z}$ indicate a 5% critical value (for $k = 9$) of 27–30, and a critical value 31–32 at level α between .01 and .025. Since the shape of empirical distributions for ψ_N^2 in simulations looks so consistently like a weighted or non-central χ^2, we can hope also to use a naive method of fitting to estimate probabilities. Following Bartlett (1978), we use the first two moments in fitting a modified Chi squared distribution to ψ_N^2. Let

$$a_N = E(\psi_N^2), \qquad b_N = \tfrac{1}{2}\operatorname{Var}(\psi_N^2).$$

Then the statistic

$$a_N \psi_N^2 / b_N$$

has mean a_N^2/b_N and variance $2a_N^2/b_N$. One approximates the distribution of $a_N\psi_N^2/b_N$ by

TABLE 3.1

An approximation of $G(c) = \lim_{N\to\infty} P(\psi_N^2 > c)$, $c = 30, 32$, *for* $k = 9$, *obtained from* 100 *observations of* ψ_{1000}^2. U_t *are* i.i.d $N(0, 1)$

	$\hat{f} = \hat{a}_N^2/\hat{b}_N$	$\hat{G}(30)$	$\hat{G}(32)$
$Z_t = U_t$	7.604	0.025	0.017
$Z_t = U_t + 0.4U_{t-1}$	7.961	0.020	0.012
$Z_t = U_t - 0.7U_{t-1}$	7.672	0.036	0.017
$Z_t = 0.25Z_{t-1} + U_t$	9.144	0.025	0.015
$Z_t = 0.8Z_{t-1} + U_t$	9.337	0.030	0.020
$Z_t = -0.36Z_{t-1} + U_t$	6.490	0.028	0.020
$Z_t = -0.72Z_{t-1} + U_t$	9.622	0.033	0.022
$Z_t = 2U_t + 85U_{t-1}$	8.875	0.027	0.018
$Z_t = \sin(2U_t + 85U_{t-1})$	8.867	0.016	0.010
$Z_t = 0.9U_t + 0.8U_{t-1} - 6.2U_{t-2} - 9.3U_{t-3} + 0.4U_{t-4}$	8.009	0.020	0.013
$Z_t = 0.8U_t - 0.9U_{t-1} + 0.3U_{t-2} + 0.1U_{t-3} - 0.5U_{t-4}$	7.535	0.040	0.025
$Z_t = 21U_t + 38U_{t-1} + 4.1U_{t-2} - 0.6U_{t-3} - 9.9U_{t-4}$	8.570	0.023	0.015
$Z_t = 20U_t - 5U_{t-1} + 30U_{t-2} - 8U_{t-3} + 75U_{t-4}$	5.863	0.045	0.035

a Chi squared distribution with a_N^2/b_N degrees of freedom. Using this method, in Table 3.1 we list approximations to the asymptotic tail probabilities of ψ_N^2 for various processes. For a given process, the mean and variance of ψ_{1000}^2 were obtained by simulating the process one hundred times, where for each simulation, ψ_{1000}^2 was obtained from nine classes. That is

$$\psi_{1000}^2 = \sum_{j=1}^{9} \frac{(\Delta_{j,1000} - m_{j,1000})^2}{m_{j,1000}}.$$

The rule of thumb with $k = 9$ appears to be that the distribution of ψ_N^2 is approximately $\Gamma(4.0, 0.29)$; for further discussion of this and other approximations, see Kedem and Slud (1981). Serious departures from this rough approximation occur for ARMA as well as nonlinear processes (e.g. the eighth process in Table 3.1 and the three processes in Table 3.2) but seem to have more to do with the abruptness of changes along sample paths than with linear or nonlinear structure per se.

A detailed exposition of the application of ψ^2 to discriminate ARMA models from unspecified alternatives is given in Kedem and Slud (1981). In particular, the effect on ψ^2 of estimating model parameters is recognized and discussed in simple examples. As another example of the application of ψ^2 for discrimination, we consider the Bernoulli(p) point-processes $\{Z_t\}$ with adjoined normal variates (cf. Grenander, 1959), defined as follows. Let $\{\varepsilon_n\}_{n=-\infty}^{\infty}$ be a Bernoulli sequence of Binomial $(1, p)$ random variables, and $\{T_i\}_{i=-\infty}^{\infty} \equiv \{n \in \mathbb{Z} : \varepsilon_n = 1\}$. If $\{M_k\}_{k=-\infty}^{\infty}$ is an i.i.d. sequence of standard normal variables, independent of $\{\varepsilon_n\}$, put $Z_t \equiv M_k$ whenever $T_k \le t < T_{k+1}$.

As in previous simulations, for $p = .25$, $.5$, and $.75$ we obtained one hundred binary arrays $\mathbf{X}_{1000}^{(9)}$. In Table 3.2 we exhibit typical higher-order crossing numbers $D_{j,1000}$ for one such array for each p.

The histograms for ψ_{1000}^2 as well as the approximate gamma tail probabilities indicate 32 as an appropriate critical value for α between .01 and .025 for the three values of p. Now we compute ψ_{1000}^2 in order to illustrate discrimination among the parameter values $p = .25$, .5, and .75. In each case the mean-values $m_{j,N}$ (which should really be known exactly) were estimated as sample averages.

We first compute ψ_N^2 with k fixed at 9 to check whether the first column in Table 3.2 indeed arises from the process with $p = .25$:

$$\psi_{1000}^2 = \frac{(125\text{–}123.66)^2}{123.66} + \frac{(95\text{–}101.7)^2}{101.7} + \frac{(156\text{–}156.41)^2}{156.41} + \cdots + \frac{(224\text{–}222.54)^2}{222.54} = 1.6 \ll 32,$$

and the null hypothesis $E(\Delta_{j,N}) = m_{j,N}$ is accepted as it should be. On the other hand, if we

TABLE 3.2

Higher order crossings for three Bernoulli(p) processes with adjoined normal variates. (Second column for each p are sample averages of 100 $D_{j,N}$ *values.)*

	p = 0.25		**p = 0.50**		**p = 0.75**	
$D_{1,1000}$	125	123.76	258	248.4	329	325.5
$D_{2,1000}$	220	225.4	419	416.4	476	492.0
$D_{3,1000}$	376	381.8	593	592.2	590	600.2
$D_{4,1000}$	503	512.7	725	715.0	663	672.9
$D_{5,1000}$	599	602.7	773	770.2	687	700.0
$D_{6,1000}$	677	678.7	805	806.5	715	719.6
$D_{7,1000}$	733	731.4	821	826.6	722	731.9
$D_{8,1000}$	775	776.5	828	841.3	732	744.2
$(\text{Ave}(\psi^2), \text{Var}(\psi^2))$	(10.1, 45.4)		(12.7, 47.6)		(13.0, 47.2)	
$\hat{f}$	4.5		6.8		7.2	
$(\hat{G}(30), \hat{G}(32))$	(.014, .010)		(.022, .015)		(.023, .015)	

compute $\psi_N^2 = \sum_{j=1}^{9} (\Delta_{j,N} - \hat{m}_{j,N})^2/\hat{m}_{j,N}$ for the $\Delta_{j,N}$ in the *second* column of Table 3.2 while $m_{j,N}$ correspond to $p = 0.25$, we find $\psi_{1000}^2 = 146 + 34.6 + \cdots + 12 \gg 32$. For $\Delta_{j,N}$ chosen from the third column of Table 3.2, with the same $m_{j,N}$ as before, $\psi_{1000}^2 = 341 + 20.2 + \cdots + 8.88 \gg 32$. Of course, discrimination among the three columns of Table 3.2 is easy, and the behavior of our statistic ψ_N^2 is agreeably decisive.

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

BARTLETT, M. (1978). *Stochastic Processes,* 3rd ed. Cambridge University Press, Cambridge, England.

BECKER, T. (1978). *Recognition of Patterns,* 3rd ed. Springer-Verlag, New York.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BOX, G. and JENKINS, G. (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

BRILLINGER, D. (1968). Estimation of the cross-spectrum of a stationary bivariate Gaussian process from its zeroes. *J. Roy. Statist. Soc. B* **30** 145–159.

COX, D. R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. B* **24** 406–424.

CRAMER, H. and LEADBETTER, M. (1968). *Stationary and Related Processes.* Wiley, New York.

DOOB, J. (1953). *Stochastic Processes.* Wiley, New York.

GERSCH, W. and BROTHERTON, T. (1979). Discrimination in locally stationary time series. *Proc. D.D.C.*, San Diego.

GERSCH, W., MARTINELLI, F., YONEMOTO, J., LOW, M. D. and MCEWAN, J. (1979). Automatic classification of electroencephalograms: Kullback-Leibler nearest neighbor rules. *Science* **205** 193–195.

GRENANDER, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.* **1** 195–277.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.

HINICH, M. (1967). Estimation of spectra after hard clippings of Gaussian processes. *Technometrics* **9** 391-400.

JOHN, E., AHN, H., PRICHEP, L. et al. (1980). Developmental equations for the electroencephalogram. *Science* **210** 1255–1258; Developmental equations reflect brain dysfunctions. *Science* **210** 1259–1262.

KEDEM, B. (1980). *Binary Time Series.* Dekker, New York.

KEDEM, B. (1981). A convergence phenomenon and nonparametric classification of random fields. Ben Gurion University Technical Report.

KEDEM, B., MENDELSON, W. B. AND GILLIN, J. C. (1981). Discrimination between REM, WAKE and NONREM sleep of the rat using higher order crossings. Unpublished report.

KEDEM, B. and SLUD, E. (1979). Higher order crossings in the discrimination of time series, I. Univ. of Maryland Tech. Report TR79-66.

KEDEM, B. and SLUD, E. (1981). On goodness of fit of time series models: An application of higher order crossings. *Biometrika* **68** 551–556.

NIEDERJOHN, R. and CASTELAZ, P. (1978). Zero-crossing analysis methods for speech recognition. I.E.E.E. International Conf. on Acoustics, Speech, and Signal Processing (CH1318), 507–513.

SCHIESS, J. R. (1979). Zero crossings counts as a method of classifying digital signals. Report, NASA Langley Research Center, Hampton, Va.

SLUD, E. (1981). Statistics of random processes with stationarity-renewal. University of Maryland technical report TR 81-35.

WALKER, A. M. (1967). Some tests of separate families of hypotheses in time series analysis. *Biometrika* **54** 39–68.

WHITTLE, P. (1951). *Hypothesis Testing in Time Series Analysis.* Thesis, Uppsala University.

MATHEMATICS DEPARTMENT
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742