Biometrika (1993), **80**, 3, *pp.* 611–22
Printed in Great Britain

Spectral analysis for categorical time series: Scaling and the spectral envelope

BY DAVID S. STOFFER
Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, U.S.A.

DAVID E. TYLER AND ANDREW J. MCDOUGALL
Department of Statistics, Rutgers University, New Brunswick, New Jersey 08903, U.S.A.

SUMMARY

Many studies produce categorical time series in which harmonic analysis is of interest. Although there exist time domain approaches for the analysis of categorical time series such as Markov chains or link function based regression models, there is apparently little statistical theory or methodology for analyzing qualitative-valued time series in the frequency domain. The purpose of this paper is to initiate the development of a general framework for the frequency domain analysis of categorical time series. In doing so, we discuss the scaling of categorical time series and introduce a new concept that we call the spectral envelope of a categorical time series. We demonstrate our methodology on a data set from a problem in molecular biology.

Some key words: Asymptotic distribution of latent roots and vectors; DNA sequencing; Frequency domain analysis; Markov chain; Multinomial time series; Scaling; Spectral envelope.

1. INTRODUCTION

There are numerous statistical techniques for analyzing continuous-valued time series in both the time and frequency domains. If a time series is discrete-valued, there is also a number of available techniques, for example, DARMA models (Lewis, 1980), INAR models (Al-Osh & Alzaid, 1987), and truncated models (Heckman, 1981) in the time domain, Fourier and Walsh–Fourier (Stoffer, 1987) analysis in the spectral domain. If the time series is categorical-valued, then there is the theory of Markov chains (Billingsley, 1961; Raftery, 1985), and the link function approach (Fahmeir & Kaufmann, 1987) for time domain analysis. But there is apparently little statistical theory or methodology for doing frequency domain analysis of categorical time series. Categorical processes, in which harmonic analysis is of interest, occur in many fields such as the medical, behavioural, epidemiological and genetic sciences; see Stoffer (1991) for some specific examples. Since a categorical process is qualitative, it is not even clear what is meant by the spectrum of a categorical process. For earlier attempts at this problem, see Lai (1978) and Tavaré & Giddings (1989).

In this paper, we introduce an approach for the spectral analysis and scaling of categorical time series. We do not attempt to solve completely the problem of spectral analysis for categorical time series; indeed, we will most likely raise as many questions

as we answer. Our intention is to provide a sound statistical basis on which to establish the analysis of categorical time series in the frequency domain.

One approach for exploring the periodic nature of a categorical process is to assign numerical values to each of the states or categories followed by a spectral analysis of the resulting discrete-valued time series. The observed spectrum will, of course, depend on the particular values chosen for the state-space. Alternatively, rather than choose arbitrary values, we propose selecting values that help emphasize any periodic feature that may exist in the categorical process.

To motivate the problem, consider the following sequence: $\{a, b, c, b, a, b, c, b, \ldots\}$. Clearly there is not just one cyclic pattern in this sequence; for example, the given sequence can also be viewed as $\{\sim b, b, \sim b, b, \sim b, b, \ldots\}$ where $\sim b$ means 'not b'. Hence, not only is the process making one cycle every four time points, it is also making one cycle every two time points. If we were to assign numerical values to the states, one particular assignment (scaling) that emphasizes the harmonic component of one cycle every four time points is $a=0$, $b=1$, $c=2$, while the scaling $a=c=0$, $b=1$, emphasizes the harmonic component of one cycle every two time points. This example makes it clear that we do not want to focus on only one scaling, but all the scalings that will identify the variety of signals that are present in a categorical process. In § 2, we introduce a new concept called the spectral envelope which can be used to help select such scalings.

Statistical issues are discussed in § 3. In § 4 we present an analysis of a data set from a problem in molecular biology. Some technical details are given in the Appendix.

2. SCALING AND THE SPECTRAL ENVELOPE

Let X_t, for $t=0, \pm 1, \pm 2, \ldots$, be a categorical-valued time series with finite state-space $\mathscr{C}=\{c_1, c_2, \ldots, c_k\}$. Assume throughout that X_t is stationary and $p_j = \mathrm{pr}\,(X_t = c_j) > 0$ for $j=1, 2, \ldots, k$. For $\beta=(\beta_1, \beta_2, \ldots, \beta_k)' \in \mathbb{R}^k$, denote by $X_t(\beta)$ the real-valued stationary time series corresponding to the scaling that assigns c_j the value β_j $(j=1, 2, \ldots, k)$. Regardless of the scaling, assume that $X_t(\beta)$ has a continuous spectral density denoted by $f(\omega; \beta)$, for $-\pi < \omega \leqslant \pi$. Our goal is to find scalings β so that the spectral density is in some sense interesting.

Although it is not the only sensible criterion for determining β, we choose to find β to maximize the power (variance) across frequencies $\omega \in (-\pi, \pi]$, relative to the total power $V(\beta) = \mathrm{var}\,\{X_t(\beta)\}$. That is, we want to maximize $r(\beta) = \sup_\omega \{f(\omega; \beta)/V(\beta)\}$, over all β not proportional to 1_k, the $k \times 1$ vector of ones. Note that $r(\beta)$ is not defined if $\beta = a1_k$ for some $a \in \mathbb{R}$ since such a scaling corresponds to assigning each category the same value a; in this case $f(\omega; \beta) \equiv 0$ and $V(\beta) = 0$. The optimality criterion $r(\beta)$ possesses the desirable property of being invariant under location and scale changes, that is, for any $a_1 \neq 0$ and $a_2 \in \mathbb{R}$, $r(a_1\beta + a_2 1_k) = r(\beta)$.

As in most scaling problems for categorical data, it is useful to represent the categories in terms of the vectors $e_1, e_2, \ldots, e_k$, where e_j represents the $k \times 1$ vector with a one in the jth row, and zeros elsewhere. We then define a k-dimensional stationary time series Y_t by $Y_t = e_j$ when $X_t = c_j$. The time series $X_t(\beta)$ can be obtained from the Y_t time series by the relationship $X_t(\beta) = \beta' Y_t$. Assume throughout that the vector process Y_t has a continuous spectral density denoted by $f(\omega)$. For each ω, $f(\omega)$ is, of course, a $k \times k$ complex-valued Hermitian matrix. The assumption that $f(\omega)$ is continuous is necessary and sufficient for ensuring that $X_t(\beta)$ has a continuous spectral density for all $\beta \in \mathbb{R}^k$. Note that the relationship $X_t(\beta) = \beta' Y_t$ implies that $f(\omega; \beta) = \beta' f(\omega) \beta$. The optimality

criterion $r(\beta)$ can thus be expressed as

$$r(\beta) = \sup_\omega \{\beta' f(\omega)\beta / \beta' V\beta\}, \tag{2·1}$$

where V is the variance-covariance matrix of Y_t.

For any fixed t, Y_t represents a single observation from a simple multinomial sampling scheme. It readily follows that $V = D - pp'$, where $p = (p_1, p_2, \ldots, p_k)'$, and D is the $k \times k$ diagonal matrix $D = \text{diag}\,(p_1, p_2, \ldots, p_k)$. Since, by assumption, $p_j > 0$ for $j = 1, 2, \ldots, k$, it follows that rank $(V) = k-1$ with the null space of V being spanned by 1_k. For any $k \times (k-1)$ full rank matrix A whose columns are linearly independent of 1_k, $A'VA$ is a $(k-1) \times (k-1)$ positive definite symmetric matrix.

Before addressing the problem of maximizing $r(\beta)$, some additional notation is needed. Denote the real part of $f(\omega)$ by $f^{\text{re}}(\omega)$ and let the matrix A be as previously defined. For $-\pi < \omega \leq \pi$, define $\lambda(\omega)$ to be the largest eigenvalue of the determinantal equation

$$|A' f^{\text{re}}(\omega) A - \lambda A' V A| = 0, \tag{2·2}$$

and let $b(\omega) \in \mathbb{R}^{k-1}$ be any corresponding eigenvector; that is,

$$A' f^{\text{re}}(\omega) A b(\omega) = \lambda(\omega) A' V A b(\omega). \tag{2·3}$$

The eigenvalue $\lambda(\omega) \geq 0$ does not depend on the choice of A since for any two such choices, say A_1 and A_2, there exists a nonsingular matrix M such that $A_1 M = A_2$. Although the eigenvector $b(\omega)$ depends on the particular choice of A, the equivalence class of scalings associated with $\beta(\omega) = Ab(\omega)$ does not depend on A. That is, the set

$$B(\omega) = \{\beta \in \mathbb{R}^k \,|\, \beta = a_1 \beta(\omega) + a_2 1_k\} \quad (a_1 \neq 0, a_2 \in \mathbb{R})$$

does not depend on A. A convenient choice of A is $A = [I_{k-1} | 0]'$, where I_{k-1} is the $(k-1) \times (k-1)$ identity matrix and 0 is the $(k-1) \times 1$ vector of zeros. For this choice, $A' f^{\text{re}}(\omega) A$ and $A'VA$ are the upper $(k-1) \times (k-1)$ blocks of $f^{\text{re}}(\omega)$ and V, respectively. This choice corresponds to setting the last component of $\beta(\omega)$ to zero.

We are now able to solve the problem of maximizing $r(\beta)$.

THEOREM 2·1. *Under the established notation and conditions, there exists* $0 \leq \omega_0 \leq \pi$ *such that* $\lambda(\omega_0) = \sup_\omega \lambda(\omega)$. *Furthermore, if* $\beta_0 = \beta(\omega_0)$ *then* $r(\beta) \leq r(\beta_0) = \lambda(\omega_0)$.

Proof. First note by the continuity property of eigenvalues that $\lambda(\omega)$ $(-\pi < \omega \leq \pi)$ is continuous since, by assumption, $f^{\text{re}}(\omega)$ is continuous. In addition, $f^{\text{re}}(\omega) = f^{\text{re}}(-\omega)$ implies $\lambda(\omega) = \lambda(-\omega)$ so that $\lambda(\omega)$ achieves its maximum on $0 \leq \omega \leq \pi$.

Now, since the imaginary part of a Hermitian matrix is skew symmetric, $\beta' f(\omega)\beta = \beta' f^{\text{re}}(\omega)\beta$ for $\beta \in \mathbb{R}^k$. Hence, for $0 \leq \omega \leq \pi$,

$$\sup_\beta r(\beta) = \sup_\beta \sup_\omega \left\{\frac{\beta' f(\omega)\beta}{\beta' V\beta}\right\} = \sup_\omega \sup_\beta \left\{\frac{\beta' f^{\text{re}}(\omega)\beta}{\beta' V\beta}\right\}, \tag{2·4}$$

where the supremum over β is for $\beta \neq a1_k$, $a \in \mathbb{R}$. For $\beta \neq a1_k$, $a \in \mathbb{R}$, there exists a $b \in \mathbb{R}^{k-1}$ with $b \neq 0$ such that $\beta = Ab + a1_k$. Recalling that $1_k' f(\omega) 1_k = 0$ and $1_k' V 1_k = 0$, we have

$$\sup_\beta \{\beta' f(\omega)\beta / \beta' V\beta\} = \sup_{b \neq 0} \{b' A' f^{\text{re}}(\omega) A b / b' A' V A b\}. \tag{2·5}$$

The supremum over $b \neq 0$ in (2·5) occurs at $b = b(\omega)$ with (2·5) itself being equal to $\lambda(\omega)$. The theorem then follows by applying this observation to (2·4). □

Remark 2·1. Alternatively, $\lambda(\omega)$ and $\beta(\omega)$ can be defined as the largest scalar $\lambda(\omega)$ such that for some $\beta(\omega) \neq a1_k$, $a \in \mathbb{R}$, $f^{re}(\omega)\beta(\omega) = \lambda(\omega)V\beta(\omega)$. This definition avoids introducing the matrix A. For computational purposes, (2·2) and (2·3) are more convenient, especially with A chosen to be $[I_{k-1}|0]'$. Note also that $\lambda(\omega)$ cannot be viewed as the largest root of the determinantal equation $|f^{re}(\omega) - \lambda V| = 0$ since this determinant is zero for any λ.

The value $\lambda(\omega)$ itself has a useful interpretation: specifically, $\lambda(\omega)d\omega$ represents the largest proportion of the total power that can be attributed to the frequencies $\omega d\omega$ for any particular scaled process $X_t(\beta)$, with the maximum being achieved by the scaling $\beta(\omega)$. Thus, beside the optimal scaling, β_0, other scalings $\beta(\omega)$ for which the corresponding value of $\lambda(\omega)$ is relatively large warrant attention. Because of its central role, we introduce the following name for $\lambda(\omega)$.

Definition 2·1. The spectral envelope of a stationary categorical time series X_t $(t = 0, \pm 1, \pm 2, \ldots)$ is defined to be $\lambda(\omega)$ $(-\pi < \omega \leq \pi)$.

The name spectral envelope is appropriate since $\lambda(\omega)$ envelopes the standardized spectrum of any scaled process. That is, given any β normalized so that $X_t(\beta)$ has total power one, $f(\omega; \beta) \leq \lambda(\omega)$ with equality if and only if β is proportional to $\beta(\omega)$.

Although the law of the process $X_t(\beta)$ for any one-to-one scaling β completely determines the law of the categorical process X_t, information is lost when one restricts attention to the spectrum of $X_t(\beta)$. Less information is lost when one considers the spectrum of Y_t. Dealing directly with the spectral density $f(\omega)$ itself is somewhat cumbersome since it is a function into the set of complex Hermitian matrices. Alternatively, one can view the spectral envelope as an easily understood, parsimonious tool for exploring the periodic nature of a categorical time series with a minimal loss of information.

3. ESTIMATING THE SPECTRAL ENVELOPE

Suppose that we observe a finite realization of the stationary categorical time series X_t, or equivalently, the multinomial process Y_t $(t = 0, 1, \ldots, T-1)$. The theory for estimating the spectral density of a multivariate, real-valued time series is well established (Brillinger, 1981, Ch. 7; Hannan, 1970, Ch. 5; Rosenblatt, 1959), and can be applied to estimating $f(\omega)$, the spectral density of Y_t. Given an estimate $f_T(\omega)$ of $f(\omega)$, an estimate of the spectral envelope and the corresponding scalings, denoted $\lambda_T(\omega)$ and $\beta_T(\omega)$, respectively, can then be defined in a manner analogous to (2·2)–(2·3).

Throughout this section we shall use the following notation. For a complex matrix B, denote the real part of B by B^{re}, the imaginary part of B by B^{im}, and the conjugate transpose of B by B^*. For any $k \times k$ matrix B, the notation $\tilde{B}$ will refer to the upper $(k-1) \times (k-1)$ block of B. Operations on $\tilde{B}$ refer to operations on the upper $(k-1) \times (k-1)$ block of B; for example, $\tilde{B}^{-1}$ denotes the inverse of $\tilde{B}$ and not the upper $(k-1) \times (k-1)$ block of B^{-1}. For a symmetric nonnegative definite matrix B, the notation $B^{\frac{1}{2}}$ refers to the unique symmetric nonnegative definite square root of B. For a $(k-1) \times 1$ vector v, the notation $(v: 0)$ refers to the $k \times 1$ vector whose first $k-1$ elements correspond to those of v, and whose last element is zero.

For simplicity and without loss of generality, we define $\lambda_T(\omega)$ to be the largest eigenvalue of $g_T^{\mathrm{re}}(\omega)$ where

$$g_T(\omega) = \tilde{V}_T^{-\frac{1}{2}} \tilde{f}_T(\omega) \tilde{V}_T^{-\frac{1}{2}}, \tag{3·1}$$

and V_T is the sample covariance matrix of Y_t $(t = 0, 1, \ldots, T-1)$. The scaling $\beta_T(\omega)$ is then defined by $\beta_T(\omega) = (b_T(\omega) : 0)$, where $\tilde{V}_T^{\frac{1}{2}} b_T(\omega)$ is the eigenvector of $g_T^{\mathrm{re}}(\omega)$ associated with the root $\lambda_T(\omega)$. Furthermore, let $b_T(\omega)$ be normalized so that $b_T'(\omega) \tilde{V}_T b_T(\omega) = 1$, and with the first nonzero entry of $\tilde{V}_T^{\frac{1}{2}} b_T(\omega)$ taken to be positive.

To allow for the application of a general theory in obtaining asymptotic distributions for the estimates of the spectral density $f(\omega)$, we assume throughout this section that Y_t is strictly stationary and that all cumulant spectra, of all orders, exist for the series Y_t. Rather than introduce excessive notation, we refer to Brillinger (1981, Assumption 2.6.1). The periodogram of Y_t is given by

$$I_T(\omega) = (2\pi T)^{-1} d_T(\omega) d_T^*(\omega) \quad (-\pi < \omega \leqslant \pi),$$

where $d_T(\omega)$ denotes the finite Fourier transform of Y_t. Whenever the spectral estimate refers to the periodogram, we shall refer to $\lambda_T(\omega)$ as the sample spectral envelope and $\beta_T(\omega)$ as the sample scalings.

The following lemma concerning the asymptotic distribution of the periodogram represents a special case of Brillinger (1981, Theorem 7.2.4). All limiting statements are taken as $T \to \infty$, and for simplicity, the distinct frequencies ω_j $(j = 1, \ldots, J)$ are assumed to be strictly between 0 and π. Let $W(p, \nu, \Sigma)$ denote the Wishart distribution of dimension p on ν degrees of freedom and with $p \times p$ covariance parameter Σ; similarly, $W_c(p, \nu, \Sigma)$ denotes the complex Wishart distribution; see Brillinger (1981, § 4.2) for details.

LEMMA 3·1. *Under the established notation and conditions,* $I_T(\omega_j)$ $(j = 1, \ldots, J)$ *converge in distribution to independent* $W_c\{k, 1, f(\omega_j)\}$ $(j = 1, \ldots, J)$.

Since V_T converges in probability to V, $g_T(\omega_j)$ $(j = 1, \ldots, J)$ are asymptotically independent $W_c\{k-1, 1, g(\omega_j)\}$ $(j = 1, \ldots, J)$, where $g(\omega) = \tilde{V}^{-\frac{1}{2}} \tilde{f}(\omega) \tilde{V}^{-\frac{1}{2}}$. Since the eigenvalues and eigenvectors are continuous functions of a matrix argument, at least almost everywhere with respect to Lebesgue measure, the asymptotic distributions of the sample spectral envelope $\lambda_T(\omega)$ and the sample scalings $\beta_T(\omega)$ follow from Lemma 3·1.

THEOREM 3·1. *Under the established notation and conditions, and for* $f_T(\omega) = I_T(\omega)$, *the collection* $\{\lambda_T(\omega_j), \beta_T(\omega_j); j = 1, \ldots, J\}$ *converges in distribution to* $\{\lambda_j, v_j; j = 1, \ldots, J\}$, *where* $v_j = (V^{-\frac{1}{2}} u_j : 0)$, *and* $\{\lambda_j, u_j\}$ $(j = 1, \ldots, J)$ *are the largest eigenvalues and eigenvectors of independent* $W_c^{\mathrm{re}}\{k-1, 1, g(\omega_j)\}$ *matrices* $(j = 1, \ldots, J)$ *with* u_j *normalized so that* $u_j' u_j = 1$ *and the first nonzero entry of* u_j *is positive.*

The above theorem gives a representation for the limiting distribution of the sample spectral envelope and sample scalings. Although the distribution of the largest root of a Wishart matrix or of a complex Wishart matrix has been well studied, we are not aware of any results on the distribution of the largest root of the real part of a complex Wishart matrix. Except for special cases, the form of the distribution of the largest root of a Wishart matrix or of a complex Wishart matrix is not tractable and contains the other roots of the matrix argument as nuisance parameters (Muirhead, 1982, § 9.7). The distribution of the largest root of the real part of a complex Wishart matrix is more problematic since the distribution $W_c^{\mathrm{re}}(p, \nu, \Sigma)$ is not Wishart itself, and depends not only on Σ^{re} but also on Σ^{im}.

A special case of fundamental importance is the case where Y_t is white noise wherein $2\pi g(\omega) = I_{k-1}$ for $-\pi < \omega \leq \pi$. In this case, the distribution of the largest root of a $W_c^{re}\{k-1, 1, g(\omega)\}$ matrix, which arises in Theorem 3·1, has a relatively simple form.

THEOREM 3·2. *Under the established notation and conditions, if Y_t is white noise then, for $f_T(\omega) = I_T(\omega)$, the collection $\{\lambda_T(\omega_j); j = 1, \ldots, J\}$ converges in distribution to $\{\lambda_j; j = 1, \ldots, J\}$, where the λ_j $(j = 1, \ldots, J)$ are independent and identically distributed with*

$$\text{pr}\,(2\pi\lambda_1 < x) = \text{pr}\,(\chi^2_{2(k-1)} < 4x) - \pi^{\frac{1}{2}} x^{(k-2)/2} \exp(-x)\,\text{pr}\,(\chi^2_k < 2x)/\Gamma\{(k-1)/2\} \quad (x > 0).$$

Proof. By Theorem 3·1, we know that λ_j $(j = 1, \ldots, J)$ are independent and identically distributed with $2\pi\lambda_1$ being distributed as the largest root of a $W_c^{re}(k-1, 1, I_{k-1})$ matrix. Now a $W_c^{re}(k-1, 1, I_{k-1})$ distribution is the same as a $W(k-1, 2, I_{k-1})/2$ distribution and can be characterized as the distribution of $ZZ'/2$ where Z is a $(k-1)\times 2$ matrix with mutually independent standard normal entries. Observe that the distribution of $2\pi\lambda_1$ is equivalent to that of the largest root of $Z'Z/2$, and that $Z'Z$ is $W(2, k-1, I_2)$. The distribution function of the largest root of a $W(2, k-1, I_2)$ matrix, which was first derived by John (1963), can then be applied to obtain the stated result. □

If the spectral estimate $f_T(\omega)$ is chosen to be the smoothed periodogram estimate

$$f_T(\omega) = (2m+1)^{-1} \sum_{s=-m}^{m} I_T(\omega + 2\pi s/T),$$

then Lemma 3·1 holds with $I_T(\omega_j)$ and $W_c\{k, 1, f(\omega_j)\}$ replaced by

$$f_T(\omega) \quad W_c\{k, 2m+1, f(\omega_j)\}/(2m+1),$$

respectively (Brillinger, 1981; Theorem 7.3.3). Consequently, Theorem 3·1 holds when adjusted analogously. Theorem 3·2 also holds when $I_T(\omega_j)$ is replaced by the estimate $f_T(\omega)$, in which case the distribution of $2\pi\lambda_1$ is that of the largest root of a $W(k-1, 4m+2, I_{k-1})/(4m+2)$ matrix. We refer the reader to Muirhead (1982, § 9.7) for a discussion of the largest root of a $W(p, \nu, I_p)$ matrix.

Finally, we consider consistent window spectral estimates. Specifically, consider a window function $W(\alpha)$ $(-\infty < \alpha < \infty)$, that is real-valued, even, of bounded variation, with

$$\int_{-\infty}^{\infty} W(\alpha)\, d\alpha = 1, \quad \int_{-\infty}^{\infty} |W(\alpha)|\, d\alpha < \infty.$$

Define

$$f_T(\omega) = (2\pi/T) \sum_{s=1}^{T-1} W_T(\omega - 2\pi s/T) I_T(2\pi s/T), \tag{3·2}$$

where

$$W_T(\alpha) = B_T^{-1} \sum_{j=-\infty}^{\infty} W\{B_T^{-1}(\alpha + 2\pi j)\}$$

and B_T is a bounded sequence of nonnegative scale parameters such that $B_T \to 0$ and $B_T T \to \infty$ as $T \to \infty$. The following limiting distribution for (3·2) is a special case of Brillinger (1981, Theorem 7.4.4). Define $\nu_T = (B_T T)^{\frac{1}{2}}\{2\pi \int_{-\infty}^{\infty} W(\alpha)^2\, d\alpha\}^{-\frac{1}{2}}$.

LEMMA 3·2. *Under the stated conditions and assumptions, for* $f_T(\omega)$ *defined by* (3·2), $\{\nu_T(f_T(\omega_j)-f(\omega_j)); j=1,\ldots,J\}$ *converges in distribution to* $\{Z_j; j=1,\ldots,J\}$ *where the* Z_j *are mutually independent* $k\times k$ *complex matrices with* (Z_j^{re}, Z_j^{im}) *having a multivariate normal distribution with mean zero and covariance structure not dependent on the window* $W(\alpha)$.

The complete covariance structure of Z_j defined in Lemma 3·2 is given by Rosenblatt (1959), and for the sake of brevity is not restated here.

Before finding the asymptotic distribution of the estimated spectral envelope and scalings, we first need the asymptotic distribution of $g_T^{re}(\omega)$. Let Z_j^{re} be as defined in Lemma 3·2. Since $T^{\frac{1}{2}}(V_T - V) = O_p(T)$, we have, for $j=1,\ldots,J$,

$$S_{j,T} = \nu_T\{g_T^{re}(\omega_j) - g^{re}(\omega_j)\} \to S_j = \tilde{V}^{-\frac{1}{2}}\tilde{Z}_j^{re}\tilde{V}^{-\frac{1}{2}} \tag{3·3}$$

jointly in distribution, with $\{S_j\}$ having mutually independent multivariate normal distributions with mean zero and covariance structure not dependent on the window $W(\alpha)$. The covariance structure of S_j is given in the Appendix. If the largest root of $g^{re}(\omega_j)$ is distinct, then the delta method can be used to argue that $\lambda_T(\omega_j)$ and $\beta_T(\omega_j)$ are jointly asymptotically normal. This follows since the maximum eigenvalue of a symmetric matrix and the corresponding eigenvector are analytic in a neighbourhood of an argument with a distinct maximum root. Let $b(\omega)$ be normalized so that $b(\omega)'\tilde{V}^{-1}b(\omega)=1$, and so that the first nonzero element of $\tilde{V}^{-\frac{1}{2}}b(\omega)$ is positive.

THEOREM 3·3. *Under the stated conditions and assumptions, and for* $f_T(\omega)$ *defined by* (3·2), *if for each* $j=1,\ldots,J$ *the largest root of* $g^{re}(\omega_j)$ *is distinct, then*

$$\{\nu_T(\lambda_T(\omega_j)-\lambda(\omega_j))/\lambda(\omega_j),\ \nu_T(\beta_T(\omega_j)-\beta(\omega_j)); j=1,\ldots,J\}$$

converges jointly in distribution to $\{z_j, (y_j:0); j=1,\ldots,J\}$ *with* z_j, y_j *being independent for* $j=1,\ldots,J$. *Furthermore, for each* $j=1,\ldots,J$, z_j *has a standard normal distribution and is independent of* y_j *which is multivariate normal with mean zero. The covariance matrix of* $\tilde{V}^{\frac{1}{2}}y_j$ *is given by*

$$\{\lambda(\omega_j)H(\omega_j)^+ g^{re}(\omega_j)H(\omega_j)^+ - a(\omega_j)a(\omega_j)'\}/2, \tag{3·4}$$

where

$$H(\omega_j) = g^{re}(\omega_j) - \lambda(\omega_j)I_{k-1}, \quad a(\omega_j) = H(\omega_j)^+ g^{im}(\omega_j)\tilde{V}^{\frac{1}{2}}b(\omega_j),$$

and $H(\omega_j)^+$ *is the Moore–Penrose inverse of* $H(\omega_j)$.

The proof of Theorem 3·3 is fairly technical and hence is reserved for the Appendix. Asymptotic normal confidence intervals and tests for $\lambda(\omega)$ can be readily constructed using Theorem 3·3. For $\beta(\omega)$, asymptotic confidence ellipsoids and chi-squared tests can also be constructed. A simpler asymptotic test statistic, however, can be constructed by replacing the term $a(\omega)$ in (3·4) by zero. Specifically, under the conditions of Theorem 3·3,

$$2\nu_T^2\{\lambda_T(\omega)^{-1}b(\omega)'\tilde{f}_T^{re}(\omega)b(\omega) + \lambda_T(\omega)b(\omega)'\tilde{V}_T\tilde{f}_T^{re}(\omega)^{-1}\tilde{V}_T b(\omega) - 2\} \tag{3·5}$$

converges in distribution to a distribution that is stochastically less than χ^2_{k-2} and stochastically greater than χ^2_{k-3}. The proof of this last statement is also given in the Appendix. Note that the test statistic (3·5) is zero if $b(\omega)$ is replaced by $b_T(\omega)$. The reader should compare (3·5) with Anderson's (1963) asymptotic test for a principal component vector.

The asymptotic distribution of $\lambda_T(\omega_j)$ and $\beta_T(\omega_j)$ is considerably more complicated whenever the largest root of $g^{\mathrm{re}}(\omega_j)$ is multiple. Rather than give a full treatment here, we refer the reader to the techniques given by Tyler (1981) and Eaton & Tyler (1991).

4. EXAMPLE AND DISCUSSION

The analysis of DNA sequences has become an important topic in statistical science; see Karlin & Macken (1991), for example. In this example we consider the data presented by Whisenant et al. (1991) which were used in a sequence analysis of a human *Y*-chromosomal DNA fragment. A DNA molecule consists of a long string of linked nucleotides, the pyrimidines, thymine (t) and cytosine (c), and the purines, adenine (a) and guanine (g), that are linked together by a backbone. It is sufficient to represent a DNA molecule by a sequence of bases on a single strand. Thus, a strand of DNA can be represented as a sequence of letters, termed base pairs (bp), from the finite alphabet $\{a, c, g, t\}$. The DNA fragment considered by Whisenant et al. (1991) is a string of length $T = 4156$ bp; the data are listed in that article and we do not reproduce the sequence here.

The sample spectral envelope of the DNA sequence is plotted in Fig. 1 where frequency is measured in cycles per bp, $\nu = \omega/2\pi$ $(0 < \nu < \frac{1}{2})$, and the sample spectral envelope is scaled by $4\pi T^{-1}$. The scaled sample spectral envelope, $4\pi T^{-1}\lambda_T(\nu)$, is the largest eigenvalue of $4\pi T^{-1} g_T^{\mathrm{re}}(\nu)$ where $g_T(\nu) = \tilde{V}_T^{-\frac{1}{2}} \tilde{I}_T(\nu) \tilde{V}_T^{-\frac{1}{2}}$. In this case, we can regard $4\pi I_T(\nu)$ as representing the power of the process associated with frequency ν, and TV_T as representing the total power across all Fourier frequencies of the process. Hence, $4\pi T^{-1}\lambda_T(\nu)$ can be interpreted as the largest proportion of the total power at frequency ν that can be obtained for any scaling of the process X_t, that is $X_t(\beta)$. The scaled sample spectral envelope can be inspected for peaks by employing Theorem 3·2 to approximate $\mathrm{pr}\{2\pi\lambda_T(\nu) > Ty/2\}$ for $0 \leqslant y \leqslant 1$, under the assumption that the process is white noise. In Fig. 1, we also plot the approximate null significances of 0·0001 ($y = 0{\cdot}60\%$) and 0·00001 ($y = 0{\cdot}71\%$) for a single a priori specified frequency ν. The null significances were chosen small in view of the problem of making simultaneous inferences about the value of the spectral envelope over more than one frequency. Bonferroni's inequality could be considered here, but this is not rigorous since the results are asymptotic and

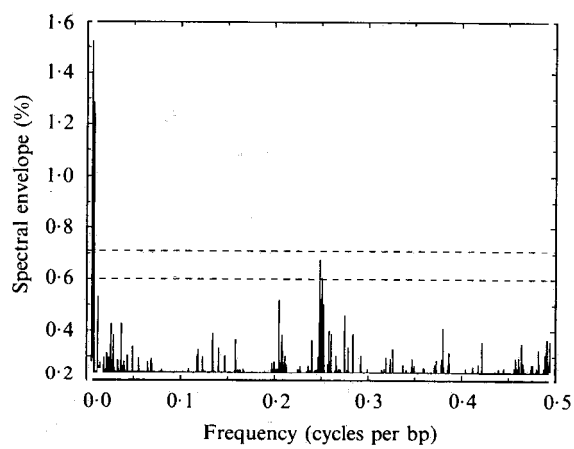

Fig. 1. Scaled sample spectral envelope of the DNA data (solid line); approximate 0·0001 and 0·00001 null significance thresholds (dashed lines).

the number of frequencies grows as the length of the series grows. A rigorous proof for justifying the use of Bonferroni's inequality would most likely involve large deviation type results and we leave that for further research. Another possibility would be to develop extensions of Fisher-type tests for hidden periodicities based on the sample spectral envelope.

To facilitate the examination of the spectral envelope plotted in Fig. 1, only values of $4\pi T^{-1}\lambda_T(\nu)$ that exceed 0·23% are plotted. Figure 1 shows a major peak at approximately $\nu = 0{\cdot}0007$ cycles per bp, about three cycles in the DNA fragment of 4156 bp, with corresponding sample scaling $a = 0$, $c = 0{\cdot}09$, $g = -0{\cdot}03$, $t = 0{\cdot}10$. This particular scaling essentially puts $a = g$ and $c = t$; this pairing is well established in the DNA literature where the binary representation of the purine-pyrimidine (ag-tc) classification is frequently studied (Tavaré & Giddings, 1989). As seen from Fig. 1, there are other large values of the sample spectral envelope in the range of frequencies $0{\cdot}0005 \leq \nu \leq 0{\cdot}003$ cycles per bp; the corresponding optimal scalings are roughly the same in this range. There is also a secondary peak at approximately $\nu = 0{\cdot}25$ cycles per bp, a period of about one cycle every four bp with a corresponding sample scaling of $a = 0$, $c = -0{\cdot}05$, $g = 0{\cdot}01$, $t = 0{\cdot}10$. Again we see the pairing of the purines, but the pyrimidines c and t are set apart; the significance of this scaling warrants further investigation and at this time we can offer no insight into this result.

Figure 2 shows the spectral envelope based on the smoothed periodogram estimate with $m = 25$. Here we see that the peaks are more well defined and occur at approximately $\nu = 0{\cdot}001$ cycles per bp with corresponding scaling $a = 0$, $c = 0{\cdot}12$, $g = -0{\cdot}01$, $t = 0{\cdot}09$, and at approximately $\nu = 0{\cdot}25$ cycles per bp with corresponding scaling $a = 0$, $c = 0{\cdot}07$, $g = 0{\cdot}01$, $t = -0{\cdot}08$. These scalings are consistent with the scalings that resulted in the periodogram based analysis shown in Fig. 1, and more strongly support the aforementioned groupings of a, g, c and t.

Next, we employed (3·5) to get asymptotic tests for the scalings $\beta_1 = \beta(0{\cdot}001) = (0, 1, 0, 1)'$, and $\beta_2 = \beta(0{\cdot}25) = (0, -1, 0, 1)'$, where the elements of β_1 and β_2 correspond to the scalings for $\{a, c, g, t\}$ in that order. The asymptotic p-value for the test of β_1 was between 0·06 and 0·16, while the asymptotic p-value for the test of β_2 was between 0·57 and 0·85. The smoothed periodogram for the scaled process $X_t(\beta_1)$, say $f_T(\nu; \beta_1)$, is displayed in Fig. 3(a) and shows a marked peak at approximately $\nu = 0{\cdot}001$; note that

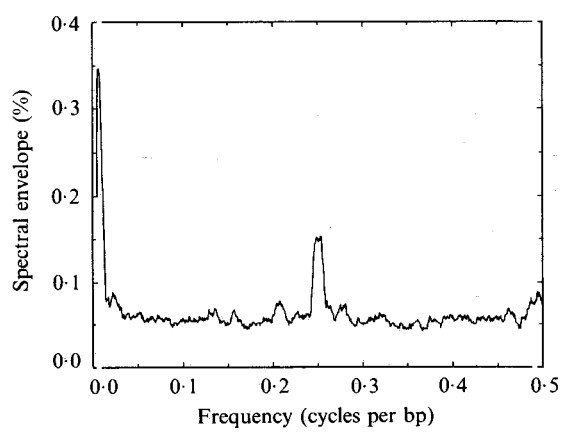

Fig. 2. Spectral envelope based on a smoothed periodogram estimate of the DNA data.

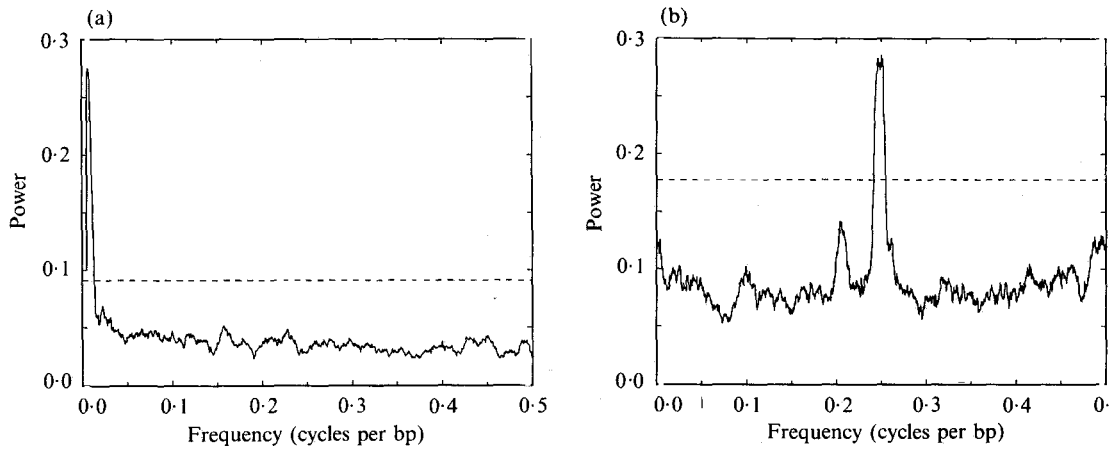

Fig. 3. Smoothed periodogram of the scaled DNA data using the scaling $a = g = 0$. (a) $c = t = 1$ (solid line); lower edge of an approximate 99·99% one-sided confidence interval for the value of the spectrum at frequency 0·001 cycles per bp (dashed line). (b) $c = -1$, $t = 1$ (solid line); lower edge of an approximate 99·99% one-sided confidence interval for the value of the spectrum at frequency 0·25 cycles per bp (dashed line).

the peak at $\nu = 0{\cdot}25$ that appears in the sample spectral envelope is suppressed. Also displayed in Fig. 3(a) is the lower limit of an approximate 99·99% chi-squared, one-sided confidence interval for $f(0{\cdot}001; \beta_1)$. We see that the lower confidence limit is roughly two times larger than the baseline power level. Similarly, $f_T(\nu; \beta_2)$, the smoothed periodogram for $X_t(\beta_2)$, is displayed in Fig. 3(b) and shows a marked peak at approximately $\nu = 0{\cdot}25$, and the peak at $\nu = 0{\cdot}001$ is suppressed. Here, the lower limit of a 99·99% one-sided confidence interval for $f(0{\cdot}25; \beta_2)$ is approximately one and a half times larger than the baseline power level.

Finally, we note that a smooth estimate of the spectral envelope is produced from a smooth estimate of $f(\nu)$ and not by smoothing the sample spectral envelope itself. This is evident from the definition of the spectral envelope since $\lambda_T(\nu_j)$ $(j = 1, \ldots, J)$ at distinct frequencies ν_j, corresponds to different scalings, $\beta_T(\nu_j)$ $(j = 1, \ldots, J)$ of the process X_t. Thus, a weighted average of the $\lambda_T(\nu_j)$ is essentially meaningless. Moreover, since the largest eigenvalue is a convex function, smoothing the sample spectral envelope would introduce positive bias into the estimate of the spectral envelope.

ACKNOWLEDGEMENTS

The work of David S. Stoffer was supported in part by grants from the National Science Foundation and the National Security Agency. The work of David E. Tyler was supported in part by a grant from the National Science Foundation. The authors thank Peter Arzberger, National Science Foundation, for suggesting the possibility of using our techniques to analyze DNA data, and Joseph Naus, Rutgers University, for directing our attention to the Whisenant et al. (1991) article.

APPENDIX

Technical details

The following notation is helpful when working with the covariance structure of random matrices. For a $p \times q$ matrix B, let vec (B) denote the pq dimensional vector obtained by stacking

the columns of B. If B is $p \times q$ and C is $c \times d$, then the Kronecker product of B and C, denoted by $B \otimes C$, is the $pc \times qd$ partitioned matrix $B \otimes C = [b_{rs}\ C]$, with $r = 1, \ldots, p$ varying over the rows, and $s = 1, \ldots, q$ varying over the columns of B. The commutation or permuted identity matrix K_{pq} is the $pq \times pq$ matrix defined by

$$K_{pq} = \sum_{r=1}^{p} \sum_{s=1}^{q} E_{rs} \otimes E'_{rs},$$

where E_{rs} is the $p \times q$ matrix with a one in the (r, s) position and zeros elsewhere. Two useful properties are the following:

$$\text{vec}\,(BQC) = (C' \otimes B)\,\text{vec}\,(Q), \quad K_{cp}(B \otimes C) = (C \otimes B)K_{dq} \tag{A·1}$$

when B is $p \times q$, C is $c \times d$, and Q is $q \times c$. See Magnus & Neudecker (1979) for details.

Rosenblatt (1959) gives the covariance structure of Z_j, as defined in Lemma 3·2, in index form. For our purposes, only the covariance structure of Z_j^{re} is needed. Using the above notation and henceforth suppressing the index j, the covariance matrix of vec (Z) is

$$(I_{kk} + K_{kk})[\{f^{\text{re}}(\omega) \otimes f^{\text{re}}(\omega)\} + \{f^{\text{im}}(\omega) \otimes f^{\text{im}}(\omega)\}]/2. \tag{A·2}$$

The covariance matrix of vec (S), where S is defined in (3·3), can be obtained from (A·2) by using the properties (A·1). For $r = k-1$, the covariance matrix of vec (S) is given by

$$(I_{rr} + K_{rr})[\{g^{\text{re}}(\omega) \otimes g^{\text{re}}(\omega)\} + \{g^{\text{im}}(\omega) \otimes g^{\text{im}}(\omega)\}]/2. \tag{A·3}$$

Proof of Theorem 3·3. Taylor series expansions for the largest eigenvalue and corresponding eigenvector of a symmetric matrix whose largest root is non-multiple (Izenman, 1975) give

$$\lambda_T(\omega) = \lambda(\omega) + \nu_T^{-1} b(\omega)' \tilde{V}^{\frac{1}{2}} S_T \tilde{V}^{\frac{1}{2}} b(\omega) + O_p(\nu_T^{-2}), \tag{A·4}$$

$$\tilde{V}_T^{\frac{1}{2}} b_T(\omega) = \tilde{V}^{\frac{1}{2}} b(\omega) - \nu_T^{-1} H(\omega)^+ S_T \tilde{V}^{\frac{1}{2}} b(\omega) + O_p(\nu_T^{-2}). \tag{A·5}$$

It follows immediately from (A·4) that $z = b(\omega)' \tilde{V}^{\frac{1}{2}} S \tilde{V}^{\frac{1}{2}} b(\omega)/\lambda(\omega)$. A representation for y follows from (A·5) since $(V_T - V) = O_p(T^{-1})$ and hence $\nu_T(\tilde{V}_T^{\frac{1}{2}} - \tilde{V}^{\frac{1}{2}}) \to 0$ in probability. This then gives $\tilde{V}^{\frac{1}{2}} y = H(\omega)^+ S \tilde{V}^{\frac{1}{2}} b(\omega)$. Thus z and y are jointly multivariate normal with mean zero. The variance of z, the covariance between z and y, and the covariance matrix of $\tilde{V}^{\frac{1}{2}} y$ can be obtained from (A·3). In obtaining the expressions given in the theorem, it is helpful to note the following. Since $g^{\text{im}}(\omega)$ is skew-symmetric, $v' g^{\text{im}}(\omega) v = 0$ for any $(k-1)$-dimensional vector v. Also, $H(\omega)^+$ commutes with $g^{\text{re}}(\omega)$, and $H(\omega)^+ \tilde{V}^{-\frac{1}{2}} b(\omega) = 0$. Finally, $(B \otimes C)(Q \otimes D) = BQ \otimes CD$ whenever the dimensions of the matrices are compatible with the operations required. □

Derivation of (3·5). Let $H_T(\omega) = g_T^{\text{re}}(\omega) - \lambda_T(\omega) I_{k-1}$. Since $V_T \to V$ and $H_T(\omega) \to H(\omega)$ in probability, we have that

$$Z_T(\omega) = 2^{\frac{1}{2}} \nu_T g_T^{\text{re}}(\omega)^{-\frac{1}{2}} H_T(\omega) \tilde{V}_T^{\frac{1}{2}} \{b_T(\omega) - b(\omega)\}/\lambda_T(\omega)^{\frac{1}{2}}$$

converges in distribution to a multivariate normal distribution with mean zero and covariance matrix $P - uu'$, where P is an idempotent matrix of rank $(k-2)$, and u is a $(k-1)$-dimensional vector with $Pu = u$, and $u'u \leqslant 1$. Thus, $Z_T(\omega)' Z_T(\omega)$ converges in distribution to $Q = \chi^2_{k-3} + (1 - u'u)\chi^2_1$, with the chi-squares being independent. Note that the distribution of Q is stochastically smaller than χ^2_{k-2} and stochastically greater than χ^2_{k-3}. Since $H_T(\omega) \tilde{V}_T^{\frac{1}{2}} b_T(\omega) = 0$, it follows from simple linear algebra that (3·5) is equal to

$$Z_T(\omega)' Z_T(\omega) + 4 b(\omega)' \{\nu_T(\tilde{V}_T - \tilde{V})\} b(\omega),$$

with the second term converging to zero in probability. □

REFERENCES

AL-OSH, M. A. & ALZAID, A. A. (1987). First-order integer-valued autoregressive process. *J. Time Series Anal.* **8**, 261–76.

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34**, 122–48.

BILLINGSLEY, P. (1961). *Statistical Inference for Markov Processes.* University of Chicago Press.

BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory*, 2nd ed. San Francisco: Holden-Day.

EATON, M. L. & TYLER, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist.* **19**, 260–71.

FAHRMEIR, L. & KAUFMANN, H. (1987). Regression models for nonstationary categorical time series. *J. Time Series Anal.* **8**, 147–60.

HANNAN, E. J. (1970). *Multiple Time Series.* New York: Wiley.

HECKMAN, J. J. (1981). Dynamic discrete probability models. In *Structural Analysis of Discrete Data with Econometric Applications*, Ed. C. F. Manski and D. McFadden, pp. 114–95. Cambridge, Mass.: MIT Press.

IZENMAN, A. J. (1975). Reduced-rank regression for multivariate linear models. *J. Mult. Anal.* **5**, 248–64.

JOHN, S. (1963). A tolerance region for multivariate normal distributions. *Sankhyā* A **25**, 363–8.

KARLIN, S. & MACKEN, C. (1991). Some statistical problems in the assessment of inhomogeneities of DNA sequence data. *J. Am. Statist. Assoc.* **86**, 27–35.

LAI, C. D. (1978). First order autoregressive Markov processes. *Stoch. Proc. Appl.* **7**, 65–72.

LEWIS, P. A. W. (1980). Simple models for positive-valued and discrete-valued time series with ARMA correlation structure. In *Multivariate Analysis V*, Ed. P. R. Krishnaiah, pp. 151–66. Amsterdam: North-Holland.

MAGNUS, J. R. & NEUDECKER, H. (1979). The communication matrix: Some properties and applications. *Ann. Statist.* **7**, 381–94.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: Wiley.

RAFTERY, A. E. (1985). A model for high-order Markov chains. *J. R. Statist. Soc.* B **47**, 528–39.

ROSENBLATT, M. (1959). Statistical analysis of stochastic processes with stationary residuals. In *Probability and Statistics*, Ed. U. Grenander, pp. 246–75. New York: Wiley.

STOFFER, D. S. (1987). Walsh–Fourier analysis of discrete-valued time series. *J. Time Series Anal.* **8**, 449–67.

STOFFER, D. S. (1991). Walsh–Fourier analysis and its statistical applications (with discussion). *J. Am. Statist. Assoc.* **86**, 461–85.

TAVARÉ, S. & GIDDINGS, B. W. (1989). Some statistical aspects of the primary structure of nucleotide sequences. In *Mathematical Methods for DNA Sequences*, Ed. M. S. Waterman, pp. 117–31. Boca Raton, Florida: CRC Press.

TYLER, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9**, 725–36.

WHISENANT, E. C., RASHEED, B. K. A., OSTRER, H. & BHATNAGAR, Y. M. (1991). Evolution and sequence analysis of a human Y-chromosomal DNA fragment. *J. Mol. Evol.* **33**, 133–41.

[*Received February* 1992. *Revised September* 1992]