Tests for Multiple Peaks in the Spectra of Categorical Time Series

by Monnie McGee

Southern Methodist University

Department of Statistical Science

P.O. Box 750332

3225 Daniel Avenue Room 144 Heroy

Dallas, Texas 75275–0332

Tests for Multiple Peaks in the Spectra of Categorical Time Series

Abstract

A Fisher's-type test for the significance of the maximum ordinate of the periodogram of a categorical time series was developed in McGee and Ensor (1998). The test was based on the Walsh–Fourier periodogram, which is more suitable for data which exhibit sharp jumps rather than smooth curves. In this paper, the results to tests of the significance of peaks in the Walsh–Fourier periodogram are extended to peaks other than the maximum.

Keywords. categorical time series, spectral envelope, Walsh–Fourier periodogram, extreme–value theory.

1 Introduction

Categorical time series are serially correlated data which are gathered in terms of states (or categories) at discrete time points. The categories respresent a nomial measure scale. Such series are found in many fields of application, including medicine (Stoffer *et al.*,1988), meteorology (Chang *et al.*,1984), and geosciences (Negi, *et a.*, 1993). A recent overview of time domain methods for the analysis of categorical time series is provided in Fokianos and Kedem (2003).

In many of these applications, it is desirable to detect a pattern in the time series and analyze its frequency characteristics. Spectral analysis is one tool by which statisticians determine the existence of a pattern in a time series. In order to decompose a time series into frequency components, a transform based on a set of complete orthogonal functions is used.

McGee and Ensor (1998) developed several tests for significance of the maximum peak in the spectra of categorical time series. The tests were based on decomposing the data with Walsh functions, a complete orthogonal set of functions which consist of square waveforms taking on the values of $+1$ and -1. Because the Walsh functions are not smooth, they are able to follow discrete levels in categorical time series (Beauchamp, 1984). Fourier analysis, based on smooth sine and cosine waves, cannot capture accurately the sharp changes found in categorical time series.

Typically, the states in a categorical time series are assigned numbers in order to facilitate graphing and analysis. However, assigning numbers to a series of states can often mask some periodic behaviours while highlighting others. Stoffer, *et al.* (1993) proposed representing a categorical time series $X(t)$ as a matrix whose dimension is the number of categories (C) by the sample size (N). A C dimensional time series is defined by $\mathbf{R}(t) = e_k$ when category k occurs at time t, where e_k is a column vector of length C with a 1 in the k^{th} row and zeros elsewhere. This solves the problem of assigning values to the categories of the time series for the case of categories representing a nominal measurement scale.

Using the above matrix–valued time series representation, the Walsh–Fourier transform at sequency λ is defined as

$$d_N(\lambda) = N^{-1/2} \sum_{t=0}^{N-1} \mathbf{R}(t) W(t, \lambda) \qquad 0 \leq \lambda < 1. \tag{1}$$

For the above equation, $\mathbf{R}(0)$, $\mathbf{R}(1)$, …, $\mathbf{R}(N-1)$ be a sample of length $N = 2^q$ from a $C \times 1$ vector–valued stationary time series $\{\mathbf{R}(t),\ t = 0, \pm 1, \pm 2, \ldots\}$. This series has a $C \times C$ autocovariance function where $\mathbf{\Gamma}(h) = cov\{\mathbf{R}(t+h), \mathbf{R}(t)\},\ h = 0, \pm 1, \pm 2, \ldots$. $W(t, \lambda)$ is the value of the Walsh transform at the time and sequency of interest, where $W(t,\cdot)$ is a $1 \times N$ vector of $+1$ and -1 that makes t zero crossings in $[0,1)$. For example, $W(0,\cdot)$ would be represented by a vector of $+1$ (no zero crossings) of length N. $W(5,\cdot)$ represents a vector in which there are 5 equally-spaced switches from $+1$ to -1 within the unit interval.

The Walsh–Fourier periodogram of the data is given by

$$I_N(\lambda) = d_N(\lambda) d_N^T(\lambda), \tag{2}$$

Note that this definition gives us a WFP which is a $C \times C$ matrix at each sequency λ; however, the rank of this matrix is one, which means that only one (nonzero) eigenvalue exists. Since this value is not dependent on the scaling of the time series, one need not worry about the proper way to scale a series such that the analysis is not affected. Stoffer, Tyler, and McDougall (1993) call this value the "spectral envelope", so named because it "envelopes the standardized spectrum of any scaled process".

We define the spectral envelope as

$$\omega(\lambda) = d_N^T(\lambda) d_N(\lambda), \tag{3}$$

which has the same eigenvalue as (2). In fact, it is itself this eigenvalue. The distribution of this value based on the hypothesis that the data are from a white noise process is χ^2_{C-1}, and, by the white noise assumption, the $\omega(\lambda_j)$ will be asymptotically independent (McGee and Ensor, 1998). In the sequel, we will denote $\omega(\lambda_j)$ as ω_j.

While some spectra may have only one peak of interest (the maximum), some may have two or more peaks, each of which is interpretable in the context of the data. Neonatal sleep state data are an example of data which have multiple peaks, only some of which have practical significance for the problem. Stoffer, *et. al.* (1988) found that the two largest peaks in the Walsh–Fourier periodogram could be interpreted in the context of the data. The largest peak represented the number of zero–crossings for the series, while the second largest represented the average number of minutes a given infant stayed in a particular sleep-state. It would be useful to have a test for the significance of all interesting peaks in categorical time series spectra. In the next section, the distributions of the k^{th} largest peaks of the WFP are given and a test statistic proposed. In Section 3, we perform a simulation study of the size and power of the test statistics. We conclude with an example of neonatal sleep state data.

2 The Distribution of the kth Largest Ordinate

Before giving the distribution of the k^{th} largest ordinate, we first review the distribution of the maximum ordinate. Further details are found in McGee and Ensor (1998) and McGee (1994).

The distribution of the maximum periodogram ordinate (W_d) involves the maximum order statistic from χ^2_{C-1} variates. From extreme–value theory, we know that the χ^2_{C-1} distribution belongs to the domain of attraction of the *exponential type* extreme–value distribution (Gumbel, 1962). This means that we want to choose constants a_N and b_N such that

$$\lim_{N\to\infty} P(Z_N < a_N + b_N x) = \exp(-e^{-x}), \qquad -\infty < x < +\infty$$

where $Z_N = \max(X_{(1)}, \ldots, X_{(N)})$ and $X_{(j)}$ denotes the jth order statistic. Galambos (1978) showed that

$$a_N = \inf\left\{x : 1 - F(x) \leq \frac{1}{N}\right\}, \tag{4}$$

and

$$b_N = [1 - F(a_N)]^{-1} \int_{a_N}^{\psi(F)} [1 - F(t)]dt \tag{5}$$

where $\psi(F) = \sup\{x : F(x) < 1\}$ and $F(\cdot)$ is the cumulative distribution function. For the WFP, a_N is the $1-(1/N)$th quantile of the χ^2_{C-1} distribution, and b_N is found by approximating the incomplete Gamma function with its asymptotic expansion. However, the larger the value of C, the greater the number of terms needed in the expansion in order to obtain a good approximation for b_N. In addition, b_N can be given in closed form only when C is odd. Equation (6) shows the closed forms for $C = 3$, $C = 5$, and $C = 7$.

$$b_N = \begin{cases} 2\mathrm{N}\mathrm{e}^{-a_N/2}, & \text{if } C = 3; \\ \mathrm{N}\mathrm{e}^{-a_N/2}(4 + a_N), & \text{if } C = 5; \\ \mathrm{N}\mathrm{e}^{-a_N/2}(6 + 2a_N + \frac{a_N}{4}), & \text{if } C = 7. \end{cases} \tag{6}$$

When C is even, McGee and Ensor used an empirical value for b_N, using the fact that a_N can be considered as the "y-intercept" of the unstandardized quantile-quantile plot of the simulated maximum values from the WFP versus the true extreme value distribution, and that b_N is the "slope" of this line. Regressing these intercepts on these slopes for various values of C gives us an equation for the relationship between C and b_N. This equation is $b_N = 1.762 + 0.081C$.

With a_N and b_N defined as in equations (4) and (5), the asymptotic distribution of the maximum ordinate of the WFP is given by

$$\frac{W_d - a_N}{b_N} \xrightarrow{d} X$$

where the cumulative distribution function of X is $F_X(x) = \exp(-e^{-x})$.

The k^{th} largest periodogram ordinate (out of n such ordinates) will be denoted by $W_{n-k+1:n}$, for $k > 1$. Its distribution can be obtained easily once the distribution of the maximum is known. David and Nagaraja (2003) gave a general result for the distribution of the k^{th} largest order statistic given the maximum order statistics. Using their formulation, the distribution of the k^{th} order statistic for the WFP is

$$F_{W_{n-k+1:n}}(x) = F_W(w) \sum_{j=0}^{k-1} \frac{[\lambda(w)]^j}{j!}, \tag{7}$$

where $\lambda(w) = -\log F_W(w)$. Therefore, for the second largest ordinate, the limiting distribution is given by

$$F_{W_{n-1:n}}(w) = exp(-e^{-w})(e^{-w} + 1) \tag{8}$$

In theory, a similar result to (8) will hold for any order statistic, such as the median, with the appropriate modification of (7). Conditions for these equations are met for distributions from the "exponential type" domain of attraction for extreme value distributions. The ordinates of the Walsh–Fourier periodogram have a χ^2 distribution, which implies that they naturally meet the conditions for these formulations to hold.

David and Nagaraja (2003) further show that the standardizing constants a_N and b_N remain the same as for the maximum order statistic. It is therefore easy to use (7) to derive the distribution of any order statistic for the Walsh–Fourier periodogram. However, if we want to solve (7) for w in order to obtain quantiles of $F_{W_{n-1:n}}(x)$ (given a vector of probabilities), we must do so numerically. For the graphs and tables that follow, (8) was solved for x using the function "nlm" in the R package.

Figure 1 compares the distribution given in (8) with the empirical distribu-

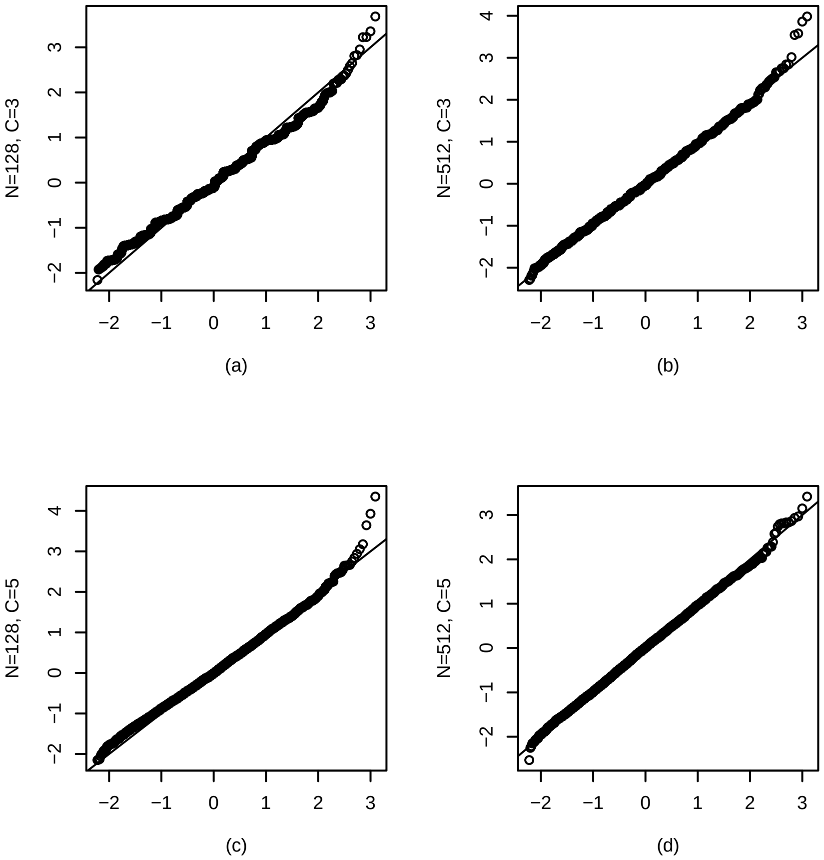


Figure 1: QQ Plots of the Empirical Distribution (on y–axis) vs. the Extreme Value Distribution (on x–axis). (a) $N = 128,\ C = 3$, (b) $N = 512,\ C = 3$, (c) $N = 128,\ C = 5$, (d) $N = 512,\ C = 5$

.

Number of Categories	*Theoretical Size*		
	0.10	0.05	0.01
$C = 2$	0.0742	0.041	0.006
$C = 3$	0.098	0.044	0.009
$C = 4$	0.098	0.049	0.010
$C = 5$	0.098	0.049	0.010

Table 1: Empirical Size of the Simulated $W_{n-1:n}$, $N = 512$.

tion based on 5000 simulated replicates of $W_{n-1:n}$ calculated from a zero-mean white noise series, where $N = 128$ and 512 and $C = 3$ and 5. Except in the right tail, the approximation is good for $N = 128$. It is good, even in the tail, for $N = 512$. The empirical size of the test for $N = 512$ and $C = 2,\ 3\ ,4,$ and 5 is given in Table 1. The sizes are based on simulations of 2000 second-largest periodogram ordinates where $N = 512$. Each simulation of 2000 was replicated 250 times. The simulation error was between 0.005 and 0.001 in all cases. Simulation error was calculated by taking the standard deviation of the alpha levels resulting from the 250 replications.

The empirical size tends to approximate well the theoretical significance level given by (8). The sizes for data with two categories are more conservative than for the others, which match the theoretical sizes very well. This is because the distribution from which we are taking the maximum is a χ^2_1. This distribution is highly skewed; therefore, we might expect slightly worse convergence properties than for the other cases. The simulated sizes for categories 4 and 5, although they seem exactly the same, are really only the same to three decimal places.

3 Scaling $W_{n-1:n}$

W_d is an example of a test statistic for testing for the significance of the maximum peak in the Walsh–Fourier periodogram. Although Fisher's test statistic is scaled, it is not necessary to scale a similar statistic based on the Walsh–Fourier periodogram. The values of the Fourier periodogram, on which Fisher's statistic is based, increase as the variance of the original time series increases. W_d is based on the spectral envelope, which is unaffected by changes in scale. In the interest of comparison, McGee and Ensor (1998) investigated three other test statistics, each one based on W_d, but divided it by another statistic calculated from the WFP (either the mean, median, or trimmed mean). They found that the trimmed mean performed best in terms of size and power for the maximum order statistic. Therefore, we examine the properties of $W_{n-1:n}$ scaled by the trimmed mean.

Since it is the large order statistics in the spectral envelope that tend to unnecessarily inflate the mean, we do not use a symmetric trimmed mean. Rather, we trim the largest r order statistics from the WFP and calculate the mean with the remaining $n - r$ ordinates. This new statistic for the test of significance of the second largest periodogram ordinate is given by

$$W_{n-1}^{r} = \frac{W_{n-1:n}}{(N-r)^{-1}\sum_{j=0}^{N-r-1}\omega_{[j]}}, \tag{9}$$

where $\omega_{[j]}$ denotes the jth order statistic of the spectral envelope. In the sequel, the term "trimmed mean" will refer to the asymmetric trimmed mean defined above. The superscript refers to the number of order statistics trimmed when calculating the mean.

It is necessary to compute the asymptotic distribution of W_{n-1}^{r}. We begin by noting that the numerator of (9) has the distribution which was derived previously. The denominator is quite a different matter. We want to find the constant, denoted μ_r, to which the denominator converges in probability.

The value of μ_r, which was originally derived in McGee and Ensor (1998), is a linear function of the order statistics, called an *L–estimate*. One form for L–estimates is given by

$$T_N = \frac{1}{N}\sum_{i=1}^{N} J\left(\frac{i}{N+1}\right) h(X_{Ni}) + \sum_{j=1}^{m} a_j X_{N,[Np_j]}, \tag{10}$$

where X_{Ni} denotes the ith order statistic out of a sample size of N; $J(u),\ 0 \leq u \leq 1$, is a generating function for weights; $h(\cdot)$ is a measurable function, $0 < p_1 < \cdots < p_m < 1$; and $a_1, \ldots, a_m$ are nonzero constants. This formulation is a linear function of the ordered values plus a weight a_j given to particular percentiles if those percentiles are a part of the L–estimate. The trimmed mean is given by (10) with $J(u) = 1/(1-\alpha)$, $0 < u < 1-\alpha$ and zero otherwise, $h(x) = x$ forevery x and $m = 0$. Here and in the sequel, $\alpha = r/N$. The limits of the sum in (10) are 1 and $N - [N\alpha]$ for the trimmed mean.

Under certain conditions stated in Chernoff, Gastwirth, and Johns (1967), and Serfling (1980), $T_n \xrightarrow{p} \mu_r$ where

$$\mu_r = \int_0^1 J(u)H(u)du + \sum_{j=1}^{m} a_j H(p_j), \tag{11}$$

and $H(\cdot) = h \circ F^{-1}$ (Serfling, 1980). The symbol "$\circ$" denotes the convolution of its operands. Since h is the identity function in the case of the trimmed mean, we need to consider only the inverse cumulative distribution function when calculating μ_r.

We perform the integration in (11) and obtain

$$\mu_r = (C-1) \cdot \frac{G_{C+1}\left[G_{C-1}^{-1}(1-\alpha)\right]}{1-\alpha}, \tag{12}$$

where $G_i(\cdot)$ is the CDF of a χ^2 random variable with i degrees of freedom. Shorack and Wellner (1988) derive the value of μ_r for general distributions. McGee (1994) gives more details for a χ^2_{C-1} distribution.

Hence, applying the previous calculations and Slutsky's theorem,

$$\frac{W^r_{n-1} - a_N}{b_N(N-r)^{-1}\sum_{j=0}^{N-r-1}\omega_{[j]}} \xrightarrow{d} \frac{X(1-\alpha)}{(C-1)G[F^{-1}(1-\alpha)]}, \tag{13}$$

where $F_X(x) = exp(-e^{-x})(e^{-x}+1)$.

Number of Categories	*Amount of Trimming*	*Theoretical Size* 0.10	0.05	0.01
2	5	0.0736	0.0410	0.0107
	26	0.0737	0.0410	0.0101
	51	0.0744	0.0410	0.0098
3	5	0.0908	0.0447	0.0097
	26	0.0911	0.0452	0.0098
	51	0.0912	0.0454	0.0098
4	5	0.0977	0.0489	0.0101
	26	0.0980	0.0490	0.0102
	51	0.0980	0.0489	0.0102
5	5	0.0978	0.0483	0.0097
	26	0.0981	0.0484	0.0098
	51	0.0981	0.0485	0.0098

Table 2: Empirical Size of the Simulated W^r_{n-1} for Various Values of r, $N = 512$.

Table 2 presents information on the empirical probability of Type I error for (9). The amount of trimming given in the table (*i.e.* 5, 26, and 51) corresponds to 1%, 5%, and 10% trimming, respectively. Again, we simulate the size for various values of the trimmed mean using 200 replications for each combination of category and amount of trimming. Two thousand maxima from a WFP were computed for each replication. Simulation errors were again between 0.001 and 0.005. Except for the size of the test for an alpha level of 0.10 and two categories, which is a bit conservative, the sizes are very close to the theoretical values.

4 Performance of Test Statistics

In this section, we explore through simulation the power of the test for the siginificance of the second largest periodogram ordinate. As an alternative model, we use the clipped AR(1) model, given by

$$X_n = \begin{cases} 1, & \text{if } Z_n \geq u; \\ 0, & \text{if } Z_n < u \end{cases} \tag{14}$$

where Z_n is an unobservable strictly stationary, continuous valued time process and u is some fixed threshold level. A clipped AR(1) series can be used as a model for which the outcome changes if a mean level crosses a certain threshold; for example, a person might sell a stock if its price goes below a certain level. The Walsh–Fourier periodogram of the clipped AR(1) model is analyzed in Stoffer and Panchalingham (1987).

For the purposes of this paper, $Z_n = \gamma_1 Z_{n-1} + \epsilon_n$, where $\gamma_1 = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ for different trials. In all trials, the threshold is 0. Only positive values for the AR(1) coefficient are examined because previous simulation revealed that the power curve is symmetric about zero. Various values of γ_1 are tried in order to ascertain if the amount of correlation in the AR(1) process affects the

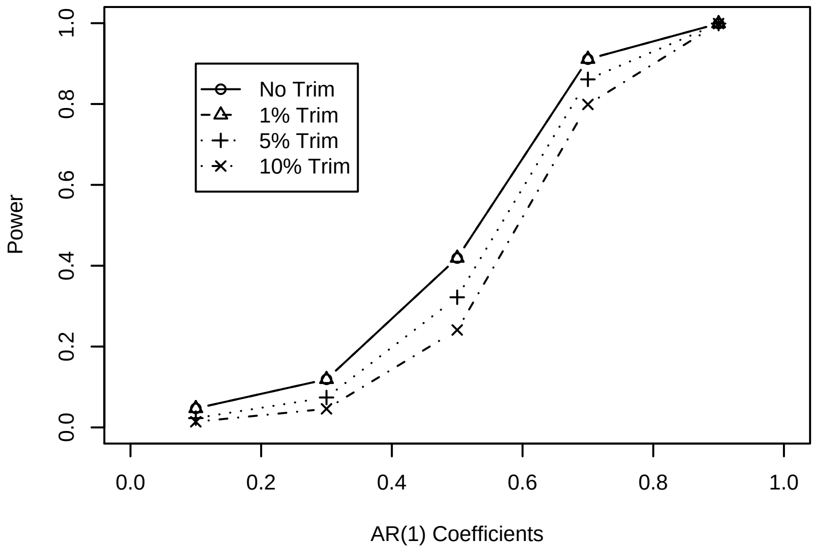


Figure 2: Coefficients of γ_1 versus Rejection Rate at the 0.05 Level of Significance for Modified Statistics.

performance of the test. We find the spectral envelope of the clipped processes and store the second largest ordinate and trimmed mean values for each run of the simulation. In all of the following simulations, $N = 512$. Before calculating the rejection rate for a given clipped AR(1) model, the results are standardized using the constants a_N given in (4) and b_N given by the empirical equation where $C = 2$.

Figure 2 is a plot of the value of γ_1, the AR(1) coefficient, versus the percent of values rejected from the clipped AR(1) model for $W_{n-1:n}$ and W^r_{n-1}. For this figure, the size of the test is 0.05. Results are similar for sizes 0.01 and 0.10, in that, the larger the correlation in the clipped AR(1) model, the higher the power. The values of r correspond to 1%, 5% and 10% trimming, in order to ascertain the effect of various amounts of trimming on the power of the test. The legend in the upper left corner of the graph shows the symbol on the graph corresponding to the plotting character for each of the modifications.

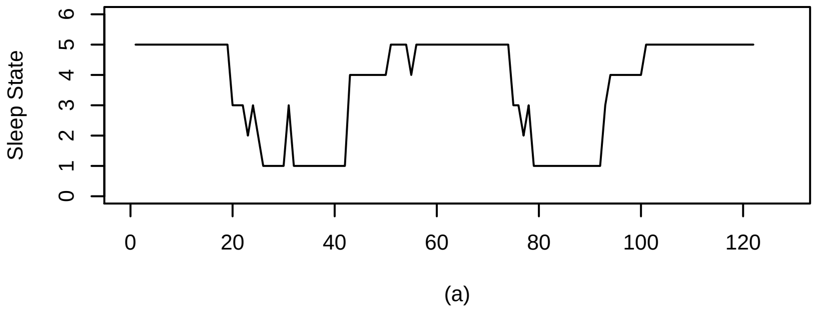
	Coefficient of AR(1) Model				
Statistic	0.1	0.3	0.5	0.7	0.9
$W_{n-1:n}$	0.047	0.119	0.419	0.911	0.999
W_{n-1}^{5}	0.047	0.119	0.419	0.911	0.999
W_{n-1}^{26}	0.023	0.074	0.322	0.861	0.999
W_{n-1}^{51}	0.014	0.046	0.241	0.799	0.999

Table 3: Coefficient of γ_1 with Corresponding Rejection Rates for $W_{n-1:n}$, and W_{n-1}^{r}, $N = 512$ and $\alpha = 0.05$.

Although it seems that the power for the unscaled test statistic and the statistic divided by the trimmed mean for $r = 5$ is the same, they are actually only the same to three decimal places. The similarity could also be caused by the fact that the peaks in the clipped AR(1) spectrum, other than the maximum, are small in magnitude. As the trimming increases, however, the gains in power decrease slightly. This would indicate that it is best to have minimal trimming, if at all. The judgment to trim should be based on the magnitude of the spurious peaks.

5 An Example

Stoffer *et. al.* (1988) investigated Walsh–Fourier analysis as applied to neonatal sleep data. The data involved 24 infants whose EEG patterns were recorded during sleep for 119 to 122 minutes each. Twelve of the infants had mothers who



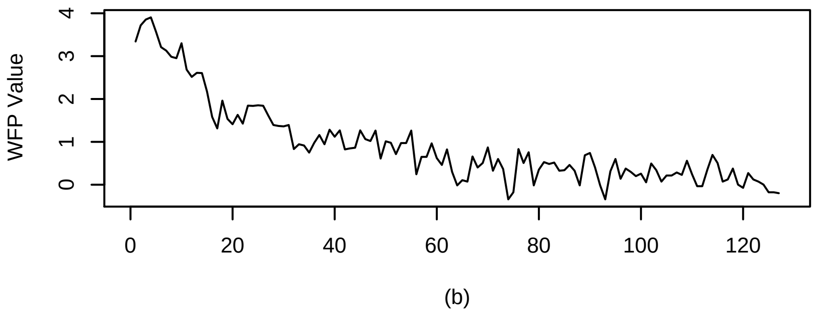


Figure 3: (a) Realization of sleep state data for one infant. (b)Average Walsh–Fourier periodogram for all 12 unexposed infant records.

had abstained from alcohol during pregnancy. The other twelve had mothers who had drank moderate amounts of alcohol during pregnancy (defined as a rate of 0.5 drinks per day). The researchers were interested in comparing the differences in sleep cycles between the two groups of neonates. We will examine the sleep cycle of the unexposed infants in order to compare the performance of W_d, $W_{n-1:n}$ and W_{n-1}^r on a real data set.

The sleep states of each infant were classified into six categories: quiet sleep - trace alternate, quiet sleep - high voltage, indeterminate sleep, active sleep -

low voltage, active sleep - mixed, and awake. The states were coded using the numbers 1 to 6, respectively.

Figure 3(a) is a plot of the sleep state of one infant for 122 minutes. An average of the WFPs for all twelve infants is given in Figure 3(b). Sleep states for the twelve infants were recorded anywhere from 113 to 122 minutes; therefore, in order to take the Walsh–Fourier transforms, the data were padded with zeros until there were 128 values for each infant. The Walsh–Fourier periodogram has been log-transformed in order to accentuate the peaks. The maximum periodogram ordinate occurs at the third sequency. This peak indicates that the data cross zero every $128/3 = 42.67$ minutes, which translates into a "period" for the data of roughly 42 minutes. Recall that this average refers to all of the infant records. For plots of all twelve infant sleep records, see Stoffer, *et. al.* (1988).

The second largest peak, occurring at the eleventh sequency, represents the average length that an infant stays in one sleep state before moving to the next. This value is $128/11 = 11.64$ minutes. Again, this is an average over all twelve infants. We wish to examine the significance of this peak, as well as that of the maximum peak.

Although the value at the eleventh sequency is the second largest peak, it is the sixth largest order statistic of the Walsh–Fourier periodogram. The test is based on the order statistics for each of the peaks of the WFP and not the peaks themselves. In other words, the second largest peak may not necessarily be the second largest order statistic for any given periodogram. For the sixth largest ordinate of the WFP, (7), for $k = 6$, becomes

$$F_{W_{n-5:n}}(x) \quad = \quad -\exp(-e^{-x})$$

$$\left[1 + e^{-x} + \frac{1}{2}e^{-2x} + \frac{1}{6}e^{-3x} + \frac{1}{24}e^{-4x} + \frac{1}{120}e^{-5x}\right] \quad (15)$$

Recall that $\lambda(w)$ is defined as $-\log\left[\exp(-e^{-x})\right] = e^{-x}$.

Table 4 gives the test statistic used, the p–value for the test, and the true size of the test statistic. The p-values are based on the theoretical asymptotic distribution of each test statistic. They are all essentially zero. The empirical sizes are simulated by calculating the percentage of time that a white noise process with $N = 128$ and $C = 6$ exceeds the ninety–fifth percentile of the extreme value distribution for both the maximum and the sixth largest order statistic (which is the second peak). Simulation results are based on 200 replications of 2000 maxima from a WFP for each of the four scenarios. Results using the trimmed test statistic are more accurate than the untrimmed statistic for the sixth ordinate; however, the accuracy does not seem to depend on the amount of trimming. The sizes are all very close to 5%, indicating that the probability of a Type I error is exactly what we would expect for this sample size and number of categories. The standard deviation of the 200 replicated size values was approximately 0.005 for all scenarios.

	Maximum Peak			Second Peak		
	Test Statistic	P-value	Size	Test Statistic	P-value	Size
$W_{n-1:n}$	49.60	< 0.001	0.056	27.12	< 0.001	0.027
W_{n-1}^{1}	10.06	< 0.001	0.049	5.50	< 0.001	0.042
W_{n-1}^{6}	14.26	< 0.001	0.050	7.80	< 0.001	0.043
W_{n-1}^{13}	19.44	< 0.001	0.050	10.63	< 0.001	0.046

Table 4: Test Statistics, P-values, and Empirical Sizes for $W_{n-1:n}$ and W_{n-1}^{r} for the Sleep State Data.

The test statistics $W_{n-1:n}$ and W^r_{n-1} have good size and power properties, and can be easily employed to examine the significance of any peak in a Walsh–Fourier periodogram. Their usage was demonstrated using neonatal sleep data. It was found that there are two significant patterns in the data - the dominate one corresponding to the sequency of the data, and the subordinate one corresponding to the amount of time an infant stays in a sleep state before moving to the next.

6 References

Beauchamp, K. G. (1984) *Applications of Walsh and Related Functions.* London: Academic Press.

Chernoff, Herman, Gastwirth, Joseph L. , and Johns, M. V. (1967) Asymptotic distributions of linear combinations of functions of order statistics. *The Annals of Mathematical Statistics*, 38, 52-72.

David, H. A. and H. N. Nagaraja (2003). *Order Statistics, 3rd Edition.* New York: John Wiley & Sons, 306-309.

Fokianos, K. and Kedem, B. J. (2003) Regression Theory for Categorical Time Series. *Statistical Science*, **18**, 357-376.

Gumbel, E. J. (1962) Statistical theory of extreme values, in *Contributions to Order Statistics*, Ahmed E. Sarhan and Bernard G. Greenburg, editors. New York: John Wiley & Sons.

McGee, Monnie (1994). *Tests for Harmonic Components in the Spectra of Categorical Time Series.* Ph. D. Dissertation, Rice University.

McGee, Monnie and K. B. Ensor. (1998) Tests for Harmonic Components in the Spectra of Categorical Time Series. *Journal of Time Series Analysis*, 19, 309 - 324.

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

Stoffer, David S. (1987) Walsh–Fourier analysis of discrete–valued time series. *Journal of Time Series Analysis*, 8, 449–467.

Stoffer, David S. (1990) Multivariate Walsh–Fourier analysis. *Journal of Time Series Analysis*, 11, 57-73.

Stoffer, David S. (1991) Walsh–Fourier analysis and its statistical applications. *Journal of the American Statistical Association*, 86, 461-485.

Stoffer, D. S, Scher, Mark S, Richardson, Gale A, Day, Nancy L, and Coble, Patricia A. (1988) A Walsh–Fourier analysis of the effects of moderate maternal alcohol consumption on neonatal sleep–state cycling. *Journal of the American Statistical Association*, 83, 954-963.