Analysis of Fisher's test for hidden periodicities

Steven J. Novick
Lehigh University

Follow this and additional works at: http://preserve.lehigh.edu/etd

Recommended Citation
AUTHOR: Novick, Steven J.

TITLE: Analysis of Fisher's Test for Hidden Periodicities

DATE: January 15, 1995
TITLE: ANALYSIS OF FISHER'S TEST FOR HIDDEN PERIODICITIES

AUTHOR: STEVEN J. NOVICK

A Thesis

Presented to the Graduate Research Committee of Lehigh University in Candidacy for the Degree of Master of Science in Applied Mathematics

Lehigh University (11/30/94)
This thesis is accepted and approved in partial fulfillment of the requirement for the Master of Science.

11/30/94
Date

Thesis Advisor

Co-Advisor

Chairperson of Department
Table of Contents

Abstract

Section 1: Modeling Periodic Data

Section 2: Grenander and Rosenblatt Approach to Fisher Statistic

Section 3: Probabilistic Approach to Fisher Statistic

Section 4: Approximate Distribution of the Fisher Statistic

Section 5: Applications

Section 6: Cross-Validation

Section 7: A Practical Guide

Section 8: Additional Questions

vita

Tables

Table 1: Frequency Distribution of p-values

Table 2: Model 2 Results

Table 3: Model 3 Results

Table 4: Model 4 Results

Table 5: Model 5 Results

Table 6: Model 6 Results

Table 7: Model 7 Results

Table 8: Cross-Validation of Model 1

Table 9: Cross-Validation of Model 4

page 1

page 2

page 7

page 17

page 24

page 27

page 34

page 37

page 39

page 49

page 29

page 30

page 31

page 32

page 32

page 33

page 34

page 36

page 36
Graphs

Graph 1A: Frequency Histogram
Graph 1B: Model 1 (Data set 55)
Graph 2: Model 2
Graph 3: Model 3
Graph 4: Model 4
Graph 5: Model 5
Graph 6: Model 6
Graph 7: Model 7

Figures

Figure 1
Figure 2
Abstract

This paper is a study of the work done on the Fisher statistic by Fisher, and later extended by Grenander and Rosenblatt. The statistic is used to test for hidden periodicities; that is, whether or not a time series contains sinusoidal components. Given data, \( x_0, x_1, \ldots, x_{n-1} \), at equally spaced time intervals, we would use the test to search for Fourier frequencies in the model.

The test is based on the \( r \)th Fisher statistic, \( G_r \), where we test \( H_0 \): there are no periodic components; that is, \( x_t = \mu + \epsilon_t \), where \( \epsilon_t \sim N(0, \sigma^2) \). The alternative is that there are at least \( r \) sinusoidal components in the model. Whenever we reject \( H_0 \), we also determine which sinusoidal components are in the model.

In this paper we provide the full details of Grenander and Rosenblatt’s derivation of the distribution of the \( r \)th Fisher statistic under the null hypothesis. Then we give a more probabilistic derivation. We then examine an approximate distribution of the \( r \)th Fisher statistic. The first approximation was given by Bloomfield. We extend Bloomfield’s result.

In analyzing some computer-generated case studies, we found that the test performed satisfactorily. Due to a phenomena called aliasing, however, we can easily reject \( H_0 \) at a particular frequency \( \omega(j) \) when the real frequency is actually \( \omega(n - j) \).

At the end of the article we discuss Shimshoni’s use of the distribution of \( G_r \) and offer an alternative way to apply it.
Section 1: Modeling Periodic Data

In nature, in the business world, and in many other aspects of our lives, we often believe that there are periodic (or cyclic) characteristics present in the data describing those phenomena. For example, we might expect that the amount of monthly rainfall in New York City increases each year until the end of April, at which time, it begins to taper off. Or, interest rates may repeatedly climb for four years and then begin to decrease again. It is usually the case, however, that we have no prior information about the periods. In this paper, we will discuss how to find these periods, if they exist, and then explore some applications.

If we believe that our data is periodic, one very simple model is to assume that there is a sinusoid embedded in noise as follows:

\[ x_t = \mu + R \cos(\omega t + \phi) + \epsilon_t, \quad t = 1, 2, \ldots, n. \]  

Here \( x_t \) denotes the \( t^{th} \) data value, \( \mu \) is the mean, \( R \) is the amplitude, \( \omega \) is the frequency (in radians per unit time), \( \phi \) is the phase, and \( \epsilon_t \) is the \( t^{th} \) error term. This paper’s focus will be to determine the frequency, \( \omega \), or the period, \( \frac{2\pi}{\omega} \).

For purposes of estimation, it is often easier to write (1) as the following:

\[ x_t = \mu + A \cos \omega t + B \sin \omega t + \epsilon_t, \]  

where \( A = R \cos \phi \) and \( B = -R \sin \phi \).

Given any values of \( A \) and \( B \), we can solve for \( R \) and \( \phi \). Thus we may consider \( \mu, A, B, \) and \( \omega \) to be our model’s parameters. As in all good-fitting models, we would like for the residuals to be as small as possible. Given \( \omega \), we may employ Least Squares on (2) to solve for \( \hat{\mu}, \hat{A}, \) and \( \hat{B} \). If we did not know \( \omega \), we may estimate
it by minimizing the residual sum of squares with respect to $\omega$. The solutions for $\hat{\mu}$, $\hat{A}$, and $\hat{B}$ are quite lengthy, and in fact, have been approximated, respectively, by

$$
\hat{\mu} = \frac{1}{n} \sum_{t=0}^{n-1} x_t = \bar{x},
$$

$$
\hat{A} = \frac{2}{n} \sum_{t=0}^{n-1} (x_t - \bar{x}) \cos \omega t,
$$

$$
\hat{B} = \frac{2}{n} \sum_{t=0}^{n-1} (x_t - \bar{x}) \sin \omega t.
$$

Not all periodic data contain just one sinusoidal component. It may also occur that the data has multiple periods. It is best to expand the model to include the $m$ periodic components:

$$
x_t = \mu + \sum_{j=1}^{m} (A_j \cos \omega_j t + B_j \sin \omega_j t) + \epsilon_t. \quad (3)
$$

While there are optimization algorithms designed to find $\hat{\mu}$, $\hat{A}_j$, and $\hat{B}_j$ (as well as estimates for $\omega_j$), the following approximate solutions have been found:

$$
\hat{\mu} = \frac{1}{n} \sum_{t=0}^{n-1} x_t = \bar{x},
$$

$$
\hat{A}_j = \frac{2}{n} \sum_{t=0}^{n-1} (x_t - \bar{x}) \cos \omega_j t,
$$

$$
\hat{B}_j = \frac{2}{n} \sum_{t=0}^{n-1} (x_t - \bar{x}) \sin \omega_j t.
$$

Note: (2) is just a special case of (3) with $m = 1$.

We will use model (3) as it is the more general, thus more flexible, model. In fact, it is often easier to use complex notation as follows:

$$
x_t = \mu + \sum_{j=1}^{m} (a_j e^{i\omega_j t}) + \epsilon_t. \quad (4)
$$
Before we begin discussing how we can find the \( \omega_j \)'s, we must examine a restriction on the values of the \( \omega_j \)'s. The fact that we are using discrete time has the effect called aliasing. Aliasing occurs when two or more frequencies become indistinguishable in the model. For example, suppose the model was simply

\[ x_t = A \cos \omega t + B \sin \omega t. \]

We would not be able to tell the difference between \( \omega \) and \( (\omega + 2\pi) \). Thus \( \omega \) and \( (\omega + 2\pi) \) are aliases. But, we also could not distinguish \( \omega \) from \( (2\pi - \omega) \) since

\[ A \cos[(2\pi - \omega)t] + B \sin[(2\pi - \omega)t] = A \cos(\omega t) - B \sin(\omega t). \]

That is, this model with parameters \((A, B, \omega)\) is identical to the same model with parameters \((A, -B, 2\pi - \omega)\). We can, therefore, restrict all frequencies to the range \( 0 \leq \omega \leq \pi \). Every frequency has an alias, called the principal alias, in that range.

We will now examine a special type of frequency called the Fourier frequency. The \( j^{th} \) Fourier frequency is \( \omega_j = \frac{2\pi j}{n} \). Because of aliasing, we need only consider frequencies \( 0 \leq \omega_j \leq \pi \), which corresponds to \( 0 \leq j \leq \frac{n}{2} \). Now suppose that \( x_0, x_1, \ldots, x_{n-1} \) are any set of \( n \) numbers. Let

\[
\begin{align*}
A_0 &= \frac{1}{n} \sum_{t=0}^{n-1} x_t = \bar{x}, \\
A_j &= \frac{2}{n} \sum_{t=0}^{n-1} x_t \cos \omega_j t, \\
B_j &= \frac{2}{n} \sum_{t=0}^{n-1} x_t \sin \omega_j t.
\end{align*}
\]
for $0 < j < \frac{n}{2}$, and if $n$ is even,

$$A_{\frac{n}{2}} = \frac{1}{n} \sum_{t=0}^{n-1} x_t \cos \omega_{\frac{n}{2}} t$$

$$= \frac{1}{n} \sum_{t=0}^{n-1} (-1)^t x_t.$$

We use Fourier frequencies in order to establish orthogonality between the sines and cosines of the Fourier frequencies (with respect to summation over the integers from 0 to $n - 1$). The orthogonality relationships are described by the following formulas (Bloomfield p 43):

$$\sum_{t=0}^{n-1} \cos \omega_j t \cos \omega_k t = \begin{cases} n/2, & \text{if } j = k \neq 0 \text{ or } n/2, \\ n, & \text{if } j = k = 0 \text{ or } n/2, \\ 0, & \text{if } j \neq k, \end{cases}$$

$$\sum_{t=0}^{n-1} \cos \omega_j t \sin \omega_k t = 0,$$

$$\sum_{t=0}^{n-1} \sin \omega_j t \sin \omega_k t = \begin{cases} n/2, & \text{if } j = k \neq 0 \text{ or } n/2, \\ 0, & \text{otherwise.} \end{cases}$$

Then from the orthogonality relations, we have

$$x_t = A_0 + \sum_{0 < j < \frac{n}{2}} (A_j \cos \omega_j t + B_j \sin \omega_j t) + (-1)^t A_{\frac{n}{2}}, \quad t = 0, 1, \ldots, n - 1, \quad (5)$$

the last term being included only if $n$ is even. Thus we can exactly represent a sequence of numbers $x_0, x_1, \ldots, x_{n-1}$ as a sum of periodic components. Some things to note:

(1) The $j^{th}$ Fourier frequency has period $\frac{2\pi}{\omega_j} = \frac{n}{j}$.

(2) This model has no error terms.

(3) Many frequencies in periodic data are not Fourier frequencies.
The effect of (3) on the $A_j$ and $B_j$ coefficients is called leakage. We will examine an example of leakage later in this paper.

Using complex notation, we can expand $x_t$ as follows:

$$x_t = \sum_{j=0}^{n-1} J_j e^{i\omega_j t}$$

where $J_j = \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-i\omega_j t}$, $j = 0, 1, \ldots, n - 1$. We say that the discrete Fourier transform of $(x_0, x_1, \ldots, x_{n-1})$ is $(J_0, J_1, \ldots, J_{n-1})$. In our analysis of the $n$ data points, we will need to take the discrete Fourier transforms. Note: $J_j = \frac{1}{2} (A_j - iB_j)$.

From here, we will only deal with Fourier frequencies. In our analysis of the frequencies, we will want to look at a special function called the periodogram, which is defined as follows:

$$I(\omega_j) = \frac{n}{2\pi} |J_j|^2 = \frac{n}{8\pi} (A_j^2 + B_j^2)$$

When scaling does not matter, $|J_j|^2$ is occasionally called the periodogram. Whatever the case, it is important to note that the periodogram is always symmetric. This is due to the fact that $|J_j|^2 = |J_{n-j}|^2$.

Suppose that $x_t = e^{i\omega t}$, for $t = 0, 1, \ldots, n - 1$. Then we would have that

$$J_j = \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-i\omega_j t} = \frac{1}{n} \sum_{t=0}^{n-1} e^{it(\omega - \omega_j)}, \text{ for some } j = 0, \ldots, n - 1$$

Clearly, $|J_j|^2$ is largest when $\omega = \omega_j$ (or if $\omega \approx \omega_j$). We would also expect that $|J_j|^2$ be considerably smaller otherwise. Thus we believe that the true periods in our data lie where the periodogram has its highest peaks; but, we cannot be certain of our results from the periodogram alone. We must test whether or not the data is truly periodic or if the data is random. That is, we would like to test the statistical significance of the periods.
In a test due to R.A. Fisher (1929), and later extended by Grenander and Rosenblatt (Grenander and Rosenblatt, pp 91-94), the null hypothesis is that no period exists; that is, \( x_t = \mu + \epsilon_t \) and that each \( \epsilon_t \) is an independent \( N(0, \sigma^2) \), for \( t = 0, 1, \ldots, n - 1 \). The alternative hypothesis (for the \( k^{th} \) Fisher statistic) would be that the model has at least \( k \) sinusoidal components. The model is given by (3) in Section 1.

We will examine the test statistic, called the Fisher statistic in three ways. First, we will examine the analytical methods employed by Grenander and Rosenblatt. Then we will re-examine the test statistic using probabilistic methods. Finally, we will explore an approximation of the distribution of the Fisher statistic found in Bloomfield (Bloomfield, pp 110-113) which will be extended in this paper.

Section 2: Grenander and Rosenblatt Approach

The methods used by Grenander and Rosenblatt can be found in their book titled *Statistical Analysis of Stationary Time Series* on pages 91-94. This paper will provide full details of Grenander and Rosenblatt’s sketchy presentation.

Suppose that \( x_0, x_1, \ldots, x_{n-1} \) are any set of \( n \) numbers. Let the number of observed values be odd, say \( n = 2m + 1 \), and consider \( m \) values of the periodogram at the points \( \omega_j = \frac{2\pi j}{2m+1}, \; j = 1, 2, \ldots, m \). Recall that the periodogram is symmetric; hence, we only need to inspect the first \( m \) points. Also, we are assuming that the sample mean was subtracted from the original data; thus, \( J_0 = 0 \).
By the null hypothesis assumption, and because of the orthogonality of the trigonometric coefficients in (5), the random variables

\[ A_j, \quad j = 1, 2, \ldots, m \]

\[ B_j, \quad j = 1, 2, \ldots, m \]

are \(2m\) independent normal variables with mean zero and variance \(\sigma^2\). Therefore each \(\frac{A_j^2 + B_j^2}{\sigma^2}\) is an independent \(\chi^2(2) = \chi^2(2)\) random variable, for \(j = 1, 2, \ldots, m\).

Let \(S_j = A_j^2 + B_j^2\), for \(j = 1, 2, \ldots, m\). The joint pdf of \(S_1, S_2, \ldots, S_m\) is

\[ f(s_1, s_2, \ldots, s_m) = c^m e^{-c \sum_i s_i}, \text{ where } c = 2/\sigma^2. \]

Let \(Y_r\) be the \(r^{th}\) largest value of \(S_1, S_2, \ldots, S_m\).

The \(r^{th}\) order Fisher statistic is defined as follows:

\[ G_r = \frac{Y_r}{Y_1 + Y_2 + \ldots + Y_m} \]

Notice that the distribution of \(G_r\) does not depend on \(\sigma^2\). In fact, without altering the value of \(G_r\), we could have defined \(Y_r\) to be the \(r^{th}\) largest value of \(|J_1|^2, |J_2|^2, \ldots, |J_m|^2\) since \(|J_j|^2 = \frac{1}{4}(A_j^2 + B_j^2)\).

We wish to find the pdf of \(G_r\) for statistical testing. What follows is the step-by-step method of Grenander and Rosenblatt's evaluation of the pdf of \(G_r\).
Let $H = \frac{1}{\sigma_r}$. Since the distribution of $H$ does not depend on the scale factor $c$, we can take it as 1. We will now find $\psi(t)$, the characteristic function of $H$. The pdf of $H$ can be found by taking the inverse Fourier transform of $\psi(t) = E[e^{itH}]$. Then to find the pdf of $G_r$, we use transformation techniques on the pdf of $H$.

Note: Below, we have that $y_1 < y_2, \ldots, y_r$ and $y_1 > y_{r+1}, \ldots, y_m$. Thus we assume that $Y_r = y_1$. There are $m$ ways to choose $Y_r$ from $\{y_1, \ldots, y_m\}$ and $\binom{m-1}{r-1}$ ways to choose the $(r - 1)$ elements that are greater than $Y_r$. Thus

$$
\psi(t) = m \binom{m-1}{r-1} \int_{y_1=0}^{y_2=y_1} \cdots \int_{y_r=y_1, y_{r+1}=0}^{y_{r+1}} \cdots \int_{y_m=0}^{\infty} e \left(\frac{\sum y_i - \sum y_j}{y_1} \right) dy_m \cdots dy_1
$$

$$
= m \binom{m-1}{r-1} \int_{y=0}^{\infty} e^{r \left(\frac{t}{1 - e^{it}}\right)} \frac{1 - e^{it}}{y^{m-r}} dy
$$

$$
= m \binom{m-1}{r-1} \frac{1}{(m-r)!} \frac{m!}{(r-1)!} \int_{y=0}^{\infty} e^{r \left(\frac{t}{1 - e^{it}}\right)} \frac{1 - e^{it}}{y^{m-r}} dy
$$

(6)

Note: This differs slightly from Grenander and Rosenblatt. They change the combinatoric $m \binom{m-1}{r-1}$ to $\frac{m!}{(m-r)! (r-1)!}$.

We will show (6) for $m = 3$ and $r = 2$. Other cases for $m = 3$ are similar. In fact, an induction proof of (6) is also similar; hence, we will not include that proof in this paper.
When \( m = 3 \) and \( r = 2 \), we have that \( y_2 > y_1 > y_3 \).

\[
\psi(t) = 3 \left( \begin{array}{c} 2 \\ 2 \end{array} \right) \int_{y_1=0}^{\infty} \int_{y_2=y_1}^{\infty} \int_{y_3=0}^{y_1} e^{it\left(\frac{y_1-y_2+y_3}{y_1}\right)} e^{-y_1-y_2-y_3} dy_3 dy_2 dy_1
\]

\[
= 3 \left( \begin{array}{c} 2 \\ 2 \end{array} \right) \int_{y_1=0}^{\infty} \int_{y_2=y_1}^{\infty} \int_{y_3=0}^{y_1} e^{it-y_1} e^{y_3} \left( \frac{it}{y_1} - 1 \right)^{-1} e^{y_3} \left( \frac{it}{y_1} - 1 \right) dy_3 dy_2 dy_1
\]

\[
= 3 \left( \begin{array}{c} 2 \\ 2 \end{array} \right) \int_{y_1=0}^{\infty} \int_{y_2=y_1}^{\infty} \left( \frac{it}{y_1} - 1 \right)^{-1} e^{it-y_1} e^{y_2} \left( \frac{it}{y_1} - 1 \right) dy_2 dy_1
\]

\[
= 3 \left( \begin{array}{c} 2 \\ 2 \end{array} \right) \int_{y_1=0}^{\infty} \left( \frac{it}{y_1} - 1 \right)^{-2} e^{it-y_1} e^{y_1} dy_1
\]

\[
= 3 \left( \begin{array}{c} 2 \\ 2 \end{array} \right) \int_{y_1=0}^{\infty} \left( \frac{it}{y_1} - 1 \right)^{-2} e^{2(it-y_1)} (1 - e^{it-y_1}) dy_1
\]

Now since the pdf of \( H \) is the inverse Fourier transform of \( \psi(t) \), we have

\[
f_H(h) = \frac{m}{2\pi} \left( \frac{m-1}{r-1} \right) \int_{t=-\infty}^{\infty} e^{-ith} \int_{y=0}^{\infty} e^{r(it-y)[1-e^{-it-y}]^{m-r}} \frac{dy}{[1-\frac{it}{y}]^{m-1}} dy dt
\]

The pdf of \( G_r \) can be found as follows:

\[
f_{G_r}(g) = f_H\left( \frac{1}{g} \right) \left| \frac{dh}{dg} \right| = \left( \frac{1}{g^2} \right) f_H\left( \frac{1}{g} \right)
\]

\[
= \frac{m}{2\pi} \left( \frac{m-1}{r-1} \right) \int_{t=-\infty}^{\infty} e^{-it\frac{1}{g}} \frac{1}{g^2} \int_{y=0}^{\infty} e^{r(it-y)[1-e^{-it-y}]^{m-r}} \frac{dy}{[1-\frac{it}{y}]^{m-1}} dy dt
\]

(7)
We would like to reverse the order of integration of (7). First we will find a bound for the absolute value of the integrand of (7). By the following equalities and inequalities, we can bound the integrand.

\[
\begin{align*}
|e^{-i\frac{1}{y}}| &= 1 \\
|e^{r(it-y)}| &= |e^{rt}e^{-y}| = e^{-y} \\
|1 - e^{it-y}| &\leq 1 + |e^{it-y}| = 1 + e^{-y} \\
|1 - \frac{it}{y}| &= \left[\left(1 - \frac{it}{y}\right)(1 + \frac{it}{y})\right]^{1/2} = \left[1 + \frac{t^2}{y^2}\right]^{1/2}
\end{align*}
\]

Thus we have the following inequality:

\[
\frac{|e^{r(it-y)}[1 - e^{it-y}]^{m-r}|}{|1 - \frac{it}{y}|^{m-1}} = \frac{|e^{r(it-y)}|\left(|1 - e^{it-y}|\right)^{m-r}}{\left(|1 - \frac{it}{y}|\right)^{m-1}} \leq \frac{e^{-y}(1 + e^{-y})^{m-r}}{(1 + \frac{t^2}{y^2})^{(m-1)/2}}
\]

Now we have that the integral of the absolute value of the integrand of (7) is bounded by the following:

\[
\int_{t=-\infty}^{\infty} \int_{y=0}^{\infty} \frac{e^{-y}(1 + e^{-y})^{m-r}}{(1 + \frac{t^2}{y^2})^{(m-1)/2}} dy \, dt
\]  

The integrand of (8) is clearly always positive. Thus we may employ Tonelli's Theorem (Royden p 309) and reverse the order of integration on (8).
Thus we have

\[
\frac{e^{-y} \int_0^\infty e^{-\frac{y}{x^2}} dx}{(1 + \frac{e^y}{y^2})^{(m-1)/2}} dt dy = \frac{e^{-y} \int_0^\infty e^{-\frac{y}{x^2}} dx}{(1 + \frac{e^y}{y^2})^{(m-1)/2}} dt dy
\]

\[
\frac{ye^{-y} \int_0^\infty e^{-\frac{y}{x^2}} dx}{(1 + x^2)^{(m-1)/2}} dx dy
\]

\[
\leq \int_{y=0}^\infty \int_{|z| \leq 1} ye^{-y} (1 + e^{-y})^{m-r} dx dy + \int_{y=0}^\infty \int_{|z| > 1} ye^{-y} (1 + e^{-y})^{m-r} \left( \int_{|z| > 1} \frac{1}{z^{m-1}} dz \right) dy
\]

\[
= 2 \int_{y=0}^\infty ye^{-y} (1 + e^{-y})^{m-r} dy + \int_{y=0}^\infty ye^{-y} (1 + e^{-y})^{m-r} \left( \int_{|z| > 1} \frac{1}{z^{m-1}} dz \right) dy
\]

The inner integrand of the second term is integrable for \( m > 2 \). And since \((1 + e^{-y})^{m-1} \leq 2^{m-1} \), we have that the outer integrand of the second term as well as the first term are also integrable. Thus (8) is integrable and bounded. Then by Fubini's Theorem (Royden p 307), we can reverse the order of integration of (7).

We will now attempt to find a closed form for \( \varphi_{\alpha}(g) \). In order to do so, we will use the following two lemmas:

**Lemma 1**

For \( \mu = 2, 3, \ldots \),

\[
\int_{-\infty}^\infty \frac{e^{itx}}{(k + it)^\mu} dt = \begin{cases} 
\frac{2\pi e^{-k\mu}}{(\mu-1)!} x^{\mu-1}, & \text{if } x > 0, \ k > 0, \\
0, & \text{if } x < 0, \ k > 0.
\end{cases}
\]

**Proof:** Let

\[
f(x) = \begin{cases} 
\frac{2\pi e^{-k\mu}}{(\mu-1)!} x^{\mu-1}, & \text{if } x > 0, \ k > 0, \\
0, & \text{if } x < 0, \ k > 0.
\end{cases}
\]
We will determine the Fourier transform of $f(z)$, as follows:

\[
\psi(t) = \int_{-\infty}^{\infty} e^{itz} \frac{2\pi e^{-kz}}{(\mu - 1)!} z^{\mu-1} dz
\]

\[
= 2\pi \int_{-\infty}^{\infty} e^{itz} \frac{(\frac{z}{2\pi})}{(\mu - 1)!} z^{\mu-1} dz
\]

\[
= \frac{2\pi}{k\mu} \int_{-\infty}^{\infty} e^{itz} \frac{(\frac{z}{2\pi})}{(\mu - 1)!} z^{\mu-1} dz
\]

\[
= \frac{2\pi}{k\mu} \left( \frac{1}{1 - \frac{1}{k}it} \right)^\mu
\]

The last step is due to the fact that we put $\psi(t)$ into the form of the characteristic function of a gamma density.

The inverse Fourier transform of $\psi(t)$ is as follows:

\[
\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itz} \frac{2\pi}{k\mu} \left( \frac{1}{1 - \frac{1}{k}it} \right)^\mu dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itz} \frac{2\pi}{k\mu} \left( \frac{1}{1 + \frac{1}{k}is} \right)^\mu ds
\]

\[
= \begin{cases} 
2\pi e^{-kz} z^{\mu-1}, & \text{if } z > 0, \ k > 0, \\
0, & \text{if } z < 0, \ k > 0.
\end{cases}
\]

This completes the proof of Lemma 1. □

Lemma 2

\[
(1 - e^{it-y})^{m-r} = \sum_{k=0}^{m-r} \binom{m-r}{k} 1^{m-r} (-1)^k (e^{it-y})^k
\]

\[
= \sum_{j=r}^{m} \binom{m-r}{j-r} (-1)^{j-r} (e^{it-y})^{j-r}
\]

Proof: This is just the binomial expansion of $(1 - e^{it-y})^{m-r}$. In the last step, we simply let $j = k + r$. □
Now by referring back to (7), we will find a closed form for \( f_{G_r}(g) \).

\[
\begin{align*}
f_{G_r}(g) &= \frac{1}{2\pi} m \left( \frac{m - 1}{r - 1} \right) \int_{t=-\infty}^{\infty} e^{-\frac{is}{g^2}} \sum_{j=r}^{m} \frac{e^{r(it-y)}}{[1 - \frac{i\pi}{y}]^{m-1}} \frac{1}{g^2} \int_{y=0}^{\infty} \frac{dy dt}{y} \int_{y=0}^{\infty} \frac{dy dt}{y} \\
&= \frac{m}{2\pi} \left( \frac{m - 1}{r - 1} \right) \int_{y=0}^{\infty} \frac{1}{g^2} \sum_{j=r}^{m} \left( -1 \right)^{j-r} y^{m-1} \int_{t=-\infty}^{\infty} e^{-\frac{ys}{g^2}} \frac{dy dt}{[y + is]^{m-1}} \\
&= \frac{m}{2\pi} \left( \frac{m - 1}{r - 1} \right) \int_{y=0}^{\infty} \frac{1}{g^2} \sum_{j=r}^{m} \left( -1 \right)^{j-r} y^{m-1} e^{-\frac{ys}{g^2}} \int_{t=-\infty}^{\infty} e^{-\frac{ys}{g^2}} \frac{dy dt}{[y + is]^{m-1}} 
\end{align*}
\]

But, by Lemma 2, we have that

\[
\begin{align*}
f_{G_r}(g) &= \frac{m}{2\pi} \left( \frac{m - 1}{r - 1} \right) \int_{y=0}^{\infty} \frac{1}{g^2} \sum_{j=r}^{m} \left( -1 \right)^{j-r} y^{m-1} e^{-\frac{ys}{g^2}} \int_{t=-\infty}^{\infty} e^{-\frac{ys}{g^2}} \frac{dy dt}{[y + is]^{m-1}} \\
&= \frac{m}{2\pi} \left( \frac{m - 1}{r - 1} \right) \int_{y=0}^{\infty} \frac{1}{g^2} \sum_{j=r}^{m} \left( -1 \right)^{j-r} y^{m-1} e^{-\frac{ys}{g^2}} \int_{t=-\infty}^{\infty} e^{-\frac{ys}{g^2}} \frac{dy dt}{[y + is]^{m-1}} 
\end{align*}
\]

Now if we let \( s = -t \),

\[
\begin{align*}
f_{G_r}(g) &= \frac{m}{2\pi} \left( \frac{m - 1}{r - 1} \right) \int_{y=0}^{\infty} \frac{1}{g^2} \sum_{j=r}^{m} \left( -1 \right)^{j-r} y^{m-1} e^{-\frac{ys}{g^2}} \int_{t=-\infty}^{\infty} e^{-\frac{ys}{g^2}} \frac{dy dt}{[y + is]^{m-1}} \\
&= \frac{m}{2\pi} \left( \frac{m - 1}{r - 1} \right) \int_{y=0}^{\infty} \frac{1}{g^2} \sum_{j=r}^{m} \left( -1 \right)^{j-r} y^{m-1} e^{-\frac{ys}{g^2}} \int_{t=-\infty}^{\infty} e^{-\frac{ys}{g^2}} \frac{dy dt}{[y + is]^{m-1}} 
\end{align*}
\]

But, by Lemma 1, if \( \left( \frac{1}{g} - j \right) > 0 \), we have that

\[
\begin{align*}
f_{G_r}(g) &= \frac{m}{2\pi} \left( \frac{m - 1}{r - 1} \right) \int_{y=0}^{\infty} \frac{1}{g^2} \sum_{j=r}^{m} \left( -1 \right)^{j-r} y^{m-1} e^{-\frac{ys}{g^2}} \int_{t=-\infty}^{\infty} e^{-\frac{ys}{g^2}} \frac{dy dt}{[y + is]^{m-1}} \\
&= \frac{m}{2\pi} \left( \frac{m - 1}{r - 1} \right) \int_{y=0}^{\infty} \frac{1}{g^2} \sum_{j=r}^{m} \left( -1 \right)^{j-r} y^{m-1} e^{-\frac{ys}{g^2}} \int_{t=-\infty}^{\infty} e^{-\frac{ys}{g^2}} \frac{dy dt}{[y + is]^{m-1}} 
\end{align*}
\]
But, if \( \frac{1}{g} - j > 0 \), then we have that \( j < \frac{1}{g} \). Thus

\[
f_{G_r}(g) = \frac{m!}{(r - 1)!(m - r)!(m - 2)!} \sum_{j=r}^{\lceil 1/g \rceil} \binom{m - r}{j - r} (-1)^{j-r} \left( \frac{1}{g} - j \right)^{m-2} (m - 1)! g^{m-2}
\]

\[
= m(m - 1) \sum_{j=r}^{\lceil 1/g \rceil} \binom{m - r}{j - r} (-1)^{j-r} \left( \frac{1}{g} - j \right)^{m-2} g^{m-2}
\]

Note: The sum contributes nothing if \( j = \frac{1}{g} \). Thus we may allow the sum to go up to \( j = \frac{1}{g} \).

We now have a closed form of \( f_{G_r}(g) \); but, as \( G_r \) will be used as a test statistic in hypothesis testing, we need to be able to find a critical region. Thus we need to compute \( \Pr(G_r > x) \). In order to compute this, we will need the following fact:

\[
\int_{1/n}^{1/r \lceil 1/g \rceil} \sum_{j=r}^{1/j} f_j(g) dg = \int_{1/n}^{1/r} f_r(g) dg + \int_{1/n}^{1/(r+1)} f_{r+1}(g) dg + \ldots + \int_{1/n}^{1/(n-1)} f_{n-1}(g) dg + \int_{1/n}^{1/n} f_n(g) dg
\]

\[
= \sum_{j=r}^{n} \int_{1/n}^{1/j} f_j(t) dt
\]

Thus

\[
\int_{x}^{1/r \lceil 1/g \rceil} \sum_{j=r}^{\lceil 1/z \rceil} f_j(g) dg = \sum_{j=r}^{\lceil 1/z \rceil} \int_{x}^{1/j} f_j(t) dt
\]

Now we can compute \( \Pr(G_r > x) \).
\[
\Pr(G_r > x) = \int \frac{1}{z} m(m-1) \left( \begin{array}{c} m-1 \\ r-1 \end{array} \right) \sum_{j=r}^{[1/z]} \left( \begin{array}{c} m-r \\ j-r \end{array} \right) (-1)^{j-r}(1-jx)^{m-2} \, dz \\
= \sum_{j=r}^{[1/z]} \int \frac{1}{z} m(m-1) \left( \begin{array}{c} m-1 \\ r-1 \end{array} \right) \left( \begin{array}{c} m-r \\ j-r \end{array} \right) (-1)^{j-r}(1-jx)^{m-2} \, dt \\
= m(m-1) \left( \begin{array}{c} m-1 \\ r-1 \end{array} \right) \sum_{j=r}^{[1/z]} \left( \begin{array}{c} m-r \\ j-r \end{array} \right) (-1)^{j-r} \int_{z}^{1} (1-jt)^{m-2} \, dt
\]

Now if we let \( u = 1 - jt \),

\[
\Pr(G_r > x) = m(m-1) \left( \begin{array}{c} m-1 \\ r-1 \end{array} \right) \sum_{j=r}^{[1/z]} \left( \begin{array}{c} m-r \\ j-r \end{array} \right) (-1)^{j-r} \int_{1-jx}^{0} \frac{1}{j} u^{m-2} \, du \\
= m \left( \begin{array}{c} m-1 \\ r-1 \end{array} \right) \sum_{j=r}^{[1/z]} \left( \begin{array}{c} m-r \\ j-r \end{array} \right) \frac{1}{j} (-1)^{j-r}(1-jx)^{m-1} \\
= m \left( \begin{array}{c} m-1 \\ r-1 \end{array} \right) \sum_{j=r}^{\infty} \left( \begin{array}{c} m-r \\ j-r \end{array} \right) \frac{1}{j} (-1)^{j-r}(1-jx)^{m-1}
\]

Note: we will use the following notation extensively in the next section.

Given any function, \( f(x) \), we define

\[
f(x)_{+} = \begin{cases} 
    f(x), & \text{if } f(x) > 0, \\
    0, & \text{if } f(x) \leq 0.
\end{cases}
\]

Remark: Grenander and Rosenblatt denote

\[
\Pr(G_r > x) = \frac{m!}{(r-1)!} \sum_{j=r}^{[1/z]} \frac{(-1)^{j-r}(1-jx)^{m-1}}{j(j-r)!(m-j)!}
\]

However, if \([1/z] > m\), using Grenander and Rosenblatt’s notation, this would allow for undefined terms. Thus, the new notation with combinatorics is preferable.
Section 3: Probabilistic Approach

Whereas Grenander and Rosenblatt use analytical tools of the Fourier transform, what follows is a more probabilistic approach to the problem.

Since each of the $|J_j|^2$, for $j = 1, 2, \ldots, m$, are independently distributed $\exp(2)$ random variables (under the null hypothesis), we may view them as interarrival times in a Poisson Process with $\lambda = 2$. Let

$$U_n = \sum_{j=1}^{n} |J_j|^2, \quad n = 1, \ldots, m$$

Figure I

$$0 \mid |J_1|^2 \mid |J_2|^2 \mid \ldots \mid |J_{m-1}|^2 \mid |J_m|^2 \mid U_m$$

We will divide everything by $s = \sum_{j=1}^{m} |J_j|^2 = U_m$ in order to change the above interval into the unit interval. Define:

$$X_i = \frac{|J_i|^2}{s} \quad \text{for} \quad i = 1, \ldots, m$$
$$V_i = \frac{U_i}{s} \quad \text{for} \quad i = 1, \ldots, m$$

A plot of the $X_i$'s and $V_i$'s would appear as follows:

Figure II

$$0 \mid X_1 \mid Y_1 \mid X_2 \mid Y_2 \mid \ldots \mid Y_{m-1} \mid X_m \mid 1 = Y_m$$

Conditioned on $s$, Poisson process theory tells us that each $V_i$, for $i = 1, 2, \ldots, m - 1$, are distributed as the order statistics of $m - 1$ independent $U(0, 1)$ random variables (Karlin and Taylor p 126). However, since this is true for all possible values $s$, each $V_i$ is distributed as above, but unconditionally. Let $G_r$ be the $r^{th}$ largest $X_i$ for $i = 1, \ldots, m$. Thus $G_r$ is the $r^{th}$ Fisher statistic. Again, we would like to know how to compute $\Pr(G_r > x)$.
We will prove by induction (and via two propositions) that

Theorem:

$$\Pr(G_r > x) = m(m - 1) \sum_{j=r}^{\infty} \binom{m-r}{j-r} \frac{1}{j} (-1)^{j-r} (1 - jx)^{m-1}.$$  

Proof: Define $E_i = \{X_i > x\}, \quad i = 1, \ldots, m.$  

For $r = 1$, we have

$$\Pr(G_1 > x) = \Pr(\max\{X_1, \ldots, X_m\} > x)$$

$$= \Pr(\bigcup_{i=1}^{m} E_i)$$

$$= \binom{m}{1} \Pr(E_1) - \binom{m}{2} \Pr(E_1 E_2) + \ldots$$

$$\ldots + (-1)^{m-1} \binom{m}{m} \Pr(E_1 E_2 \cdots E_m)$$

$$= \sum_{i=1}^{m} (-1)^{i-1} \binom{m}{i} \Pr(E_1 \cdots E_i).$$

Given $X_1, \ldots, X_m$, we need to evaluate

$$\Pr_m(E_1 \cdots E_k) = \Pr_m(X_1 > x, \ldots, X_k > x), \quad \text{for } k = 1, \ldots, m.$$  

Proposition 1

$$\Pr_m(X_1 > x, \ldots, X_k > x) = (1 - kx)^{m-1}$$

Proof:

We will use induction again to prove Proposition 1.

For $k = 1$, we have

$$\Pr_m(X_1 > x) = \Pr(\min\{V_1, \ldots, V_{m-1}\} > x) = (1 - x)^{m-1}$$
Note: By symmetry, we have that $\text{Pr}_m(X_j > x) = (1 - x)^{m-1}_+$. Thus we have that
\[ f_{X_j}(x) = (m - 1)(1 - x)^{m-2}_+ , \quad j = 1, \ldots, m - 1 \]

Assume
\[ \text{Pr}_m(X_1 > x, \ldots, X_k > x) = (1 - kx)^{m-1}_+ \]

Then
\[ \text{Pr}_m(X_1 > x, \ldots, X_{k+1} > x) \]
\[ = \int_{k}^{1} \text{Pr}_m(X_2 > x, \ldots, X_{k+1} > x | X_1 = s) f_{X_1}(s) ds \]
\[ = \int_{k}^{1} \text{Pr}_{m-1} \left( X_2 > \frac{x}{1-s}, \ldots, X_{k+1} > \frac{x}{1-s} \right) (m - 1)(1-s)^{m-2}_+ ds \]
\[ = \int_{k}^{1} \text{Pr}_{m-1} \left( X_1 > \frac{x}{1-s}, \ldots, X_k > \frac{x}{1-s} \right) (m - 1)(1-s)^{m-2}_+ ds \]
\[ = \int_{k}^{1} \left( 1 - k \frac{x}{1-s} \right)^{m-2}_+ (m - 1)(1-s)^{m-2}_+ ds \]

Now by the inductive hypothesis,
\[ \text{Pr}_m(X_1 > x, \ldots, X_{k+1} > x) \]
\[ = \int_{k}^{1} \left( \frac{1-s-kx}{1-s} \right)^{m-2}_+ (m - 1)(1-s)^{m-2}_+ ds \]
\[ = \int_{k}^{1} (1-s-kx)^{m-2}_+ (m - 1) ds \]
\[ = (1 - (k+1)x)^{m-1}_+ \]

Thus we have proven Proposition 1. \(\blacksquare\)
We now have
\[
Pr(G_1 > x) = \sum_{k=1}^{\infty} (-1)^{k-1} \binom{m}{k} (1 - kx)^{m-1}
\]
\[
= m \sum_{k=1}^{\infty} \binom{m-1}{k-1} \frac{1}{k} (-1)^{k-1} (1 - kx)^{m-1}
\]

Thus we have proven the Theorem for $G_1$. We may now assume that
\[
Pr(G_k > x) = m \binom{m-1}{r-1} \sum_{j=k}^{\infty} \binom{m-r}{j-r} \frac{1}{j} (-1)^{j-r} (1 - jx)^{m-1}
\]

Then for $r = k + 1$, we have
\[
Pr(G_{k+1} > x)
\]
\[
= Pr(at \ least \ k+1 \ X's \ > \ x)
\]
\[
= Pr(at \ least \ k \ X's \ > \ x) - Pr(exactly \ k \ X's \ > \ x)
\]
\[
= Pr(G_k > x) - \binom{m}{k} Pr_m(X_1 > x, \ldots, X_k > x, \ X_{k+1} < x, \ldots, X_m < x).
\]

Therefore to complete the proof of the Theorem, we must find
\[
Pr(X_1 > x, \ldots, X_k > x, X_{k+1} < x, \ldots, X_m < x).
\]

Proposition 2

\[
Pr_m(X_1 > x, \ldots, X_k > x, X_{k+1} < x, \ldots, X_m < x)
\]
\[
= \sum_{i=k}^{\infty} \binom{m-k}{i-k} (-1)^{i-k} (1 - ix)^{m-1}
\]

Proof: Again, we will employ mathematical induction to prove Proposition 2.

Using Proposition 1, we have
\[
P_m(\bigcup_{i=1}^{k} [X_i > t]) = \sum_{i=1}^{\infty} (-1)^{i-1} \binom{k}{i} (1 - it)^{m-1}
\]
Thus

\[
Pr_m(X_1 > x, X_2 < x, \ldots, X_m < x)
= \int_x^1 Pr_m (X_2 < x, \ldots, X_m < x | X_1 = s) f_{X_1}(s) ds
= \int_x^1 Pr_{m-1} \left( X_2 < \frac{x}{1-s}, \ldots, X_m < \frac{x}{1-s} \right) (m-1)(1-s)^{m-2} ds
= \int_x^1 Pr_{m-1} \left( X_1 < \frac{x}{1-s}, \ldots, X_{m-1} < \frac{x}{1-s} \right) (m-1)(1-s)^{m-2} ds
= \int_x^1 \left[ 1 - Pr_{m-1} \left( \bigcup_{i=1}^{m-1} \left[ X_i > \frac{x}{1-s} \right] \right) \right] (m-1)(1-s)^{m-2} ds
= \int_x^1 \left[ 1 - \sum_{j=1}^{\infty} (-1)^{j-1} \binom{m-1}{j} \left( 1 - j \frac{x}{1-s} \right)^{m-2} \right] (m-1)(1-s)^{m-2} ds
= \int_x^1 (m-1)(1-s)^{m-2} ds
\]

\[
- \sum_{j=1}^{\infty} (-1)^{j-1} \binom{m-1}{j} \left( 1 - j \frac{x}{1-s} \right)^{m-2} (m-1)(1-s)^{m-2} ds
= (1-x)^{m-1} - \sum_{j=1}^{\infty} (-1)^{j-1} \binom{m-1}{j} \left( 1 - j \frac{x}{1-s} \right)^{m-2} (m-1)(1-s)^{m-2} ds
\]
But, \((1 - j \frac{x}{1-s})^m > 0\) whenever \(s < 1 - jx\). Thus we have that

\[
\Pr_m(X_1 > x, X_2 < x, \ldots, X_m < x) = (1 - x)^{m-1}_+ - \int_0^x \sum_{j=1}^{\infty} (-1)^{j-1} \binom{m-1}{j} \left(1 - j \frac{x}{1-s}\right)^{m-2}_+ (m-1)(1-s)^{m-1}_+ ds
\]

\[
= (1 - x)^{m-1}_+ - \sum_{j=1}^{\infty} (-1)^{j-1} \binom{m-1}{j} \int_0^x \left(1 - \frac{s - jx}{1-s}\right)^{m-2}_+ (m-1)(1-s)^{m-2}_+ ds
\]

\[
= (1 - x)^{m-1}_+ - \sum_{j=1}^{\infty} (-1)^{j-1} \binom{m-1}{j} (1 - (j+1)x)^{m-1}_+
\]

\[
\left(1 - x\right)^{m-1}_+ - \sum_{j=0}^{\infty} (-1)^j \binom{m-1}{j} (1 - (j+1)x)^{m-1}_+
\]

Now assume for \(k > 1\) that

\[
\Pr_m(X_1 > x, \ldots, X_k > x, X_{k+1} < x, \ldots, X_m < x)
\]

\[
= \sum_{i=k}^{\infty} \binom{m-k}{i-k} (-1)^{i-k} (1 - iz)^{m-1}_+
\]

22
Then we have

\[
\Pr_m(X_1 > z, \ldots, X_k > z, X_{k+2} < z, \ldots, X_m < z)
\]

\[
= \int_{z}^{1-kx} \Pr_m(X_2 > z, \ldots, X_{k+1} > z, X_{k+2} < z, \ldots, X_m < z \mid X_1 = s) f_X(s) \, ds
\]

\[
= \int_{z}^{1-kx} \Pr_{m-1}(X_1 > \frac{x}{1-s}, \ldots, X_{k+1} > \frac{x}{1-s}, X_{k+2} < \frac{x}{1-s}, \ldots, X_m < \frac{x}{1-s}) \ldots (m-1)(1-s)^{m-2} \, ds
\]

\[
= \int_{z}^{1-kx} \Pr_{m-1}(X_1 > \frac{x}{1-s}, \ldots, X_k > \frac{x}{1-s}, X_{k+1} < \frac{x}{1-s}, \ldots, X_{m-1} < \frac{x}{1-s}) \ldots (m-1)(1-s)^{m-2} \, ds
\]

Now by the inductive hypothesis,

\[
\Pr_m(X_1 > z, \ldots, X_k > z, X_{k+2} < z, \ldots, X_m < z)
\]

\[
= \int_{z}^{1-kx} \sum_{i=k}^{\infty} (-1)^{i-k} \binom{m-k-1}{i-k} \left(1 - i \frac{x}{1-s}\right)^{m-2} (m-1)(1-s)^{m-2} \, ds
\]

\[
= \sum_{i=k}^{\infty} (-1)^{i-k} \binom{m-k-1}{i-k} \int_{z}^{1} \left(1 - s - is\right)^{m-2} (m-1)(1-s)^{m-2} \, ds
\]

\[
= \sum_{i=k}^{\infty} (-1)^{i-k} \binom{m-k-1}{i-k} (1 - (i+1)x)^{m-1}
\]

\[
= \sum_{i=k+1}^{\infty} (-1)^{i-k-1} \binom{m-k-1}{i-k-1} (1 - ix)^{m-1}
\]

Thus we have proven Proposition 2. \(\blacksquare\)
We may now complete the proof of the Theorem with \( r = k + 1 \) as follows:

\[
\Pr(G_{k+1} > x) = \Pr(G_k > x) - \binom{m}{k} \Pr(X_1 > x, \ldots, X_k > x, X_{k+1} < x, \ldots, X_m < x)
\]

\[
= \binom{m}{k} \left( \sum_{i=k}^{\infty} \binom{m-k}{i-k} \frac{1}{i} (-1)^{i-k} (1 - ix)^{m-1} \right)
- \binom{m}{k} \left( \sum_{i=k}^{\infty} (-1)^{i-k} \binom{m-k}{i-k} (1 - ix)^{m-1} \right)
\]

\[
= \binom{m}{k} \sum_{i=k+1}^{\infty} \binom{m-k}{i-k} \frac{1}{i} (-1)^{i-k-1} (1 - ix)^{m-1} (i-k)
\]

\[
= \binom{m}{k} \sum_{i=k+1}^{\infty} \binom{m-k}{i-k-1} \frac{1}{i} (-1)^{i-k-1} (1 - ix)^{m-1}
\]

Thus by induction, we have shown that

\[
\Pr(G_r > x) = \binom{m}{r-1} \sum_{j=r}^{\infty} \binom{m-r}{j-r} \frac{1}{j} (-1)^{j-r} (1 - jx)^{m-1}
\]

This gives a probabilistic proof of the Grenander and Rosenblatt formula.

Section 4: Approximate Distribution of the Fisher Statistic

Bloomfield introduces an approximate distribution of \( G_1 \), the first Fisher statistic (Bloomfield p 140). In this paper, we will calculate approximate distributions of \( G_r \), the \( r \)th Fisher statistic, in a manner similar to Bloomfield.

Let \( I_1, \ldots, I_m \) be independent exp(1) random variables. Let \( X_r \) be the \( r \)th largest element of \( \{I_1, \ldots, I_m\} \) for \( r = 1, \ldots, m \).

We will show by induction that

\[
\Pr(X_r \leq x) = \sum_{i=0}^{r-1} \binom{m}{i} (1 - e^{-x})^{m-i} (e^{-x})^i
\]

(9)
For $r = 1$, we have

$$\Pr(X_1 \leq x) = \Pr(I_1 \leq x, \ldots, I_m \leq x) = (1 - e^{-x})^m$$

Now we will assume that

$$\Pr(X_k \leq x) = \sum_{i=0}^{k-1} \binom{m}{i} (1 - e^{-x})^{m-i} (e^{-x})^i$$

Then for $r = k + 1$,

$$\Pr(X_{k+1} \leq x)$$

$$= \Pr(X_{k+1} \leq x, X_k \leq x) + \Pr(X_{k+1} \leq x, X_k > x)$$

$$= \Pr(X_k \leq x) + \Pr(X_1 > x, \ldots, X_k > x, X_{k+1} \leq x, \ldots, X_m \leq x)$$

$$= \Pr(X_k \leq x) + \binom{m}{k} \Pr(I_1 > x, \ldots, I_k > x, I_{k+1} \leq x, \ldots, I_m \leq x)$$

$$= \sum_{i=0}^{k-1} \binom{m}{i} (1 - e^{-x})^{m-i} (e^{-x})^i + \binom{m}{k} (1 - e^{-x})^{m-k} (e^{-x})^k$$

$$= \sum_{i=0}^{k} \binom{m}{i} (1 - e^{-x})^{m-i} (e^{-x})^i$$

Thus we have shown that (9) is true.

Let $Y_m = I_1 + \ldots + I_m$.

Then we can write that the Fisher statistic, $G_r = X_r/Y_m$.

But, we also have that

$$\frac{Y_m}{m} = \frac{I_1 + \ldots + I_m}{m} = \bar{I}$$

And by the Law of Large Numbers, we have

$$\frac{Y_m}{m} \to 1 \quad \text{as } m \to \infty$$

25
Thus $Y_m \sim m$ for large $m$.

Therefore

$$mG_r = \frac{X_r}{Y_m} = \frac{m}{Y_m} \rightarrow X_r.$$

Therefore we can write

$$\Pr(mG_r \leq z) \approx \sum_{i=0}^{r-1} \binom{m}{i} (1 - e^{-z})^{m-i} (e^{-z})^i \quad (7)$$

For $r = 1$, we can derive the special approximation given by Bloomfield. We find that

$$\Pr(mG_1 \leq z) = (1 - e^{-z})^m$$

But, $\Pr(mG_1 \leq z + \log m) \approx \left(1 - \frac{e^{-z}}{m}\right)^m \approx \exp(-e^{-z})$, for large $m$.

**Proposition 3**

$$\Pr(mG_r \leq z + \log m) \approx \sum_{i=0}^{r} \exp(-e^{-z}) \frac{[e^{-z}]^i}{i!}$$

where $G_r$ is the $r^{th}$ Fisher statistic.

**Proof:**

We have already shown Proposition 3 true for $r = 1$. Now we may assume the proposition true for $r = k$ and prove it for $r = k + 1$ as follows:

$$\Pr(mG_{k+1} \leq z) \approx \sum_{i=0}^{k} \binom{m}{i} (1 - e^{-z})^{m-i} (e^{-z})^i$$

$$= \sum_{i=0}^{k-1} \binom{m}{i} (1 - e^{-z})^{m-i} (e^{-z})^i + \binom{m}{k} (1 - e^{-z})^{m-k} (e^{-z})^k$$

$$\approx \Pr(mG_k \leq z) + \binom{m}{k} (1 - e^{-z})^{m-k} (e^{-z})^k$$

26
As we did before, we can write

\[
Pr(mG_{k+1} \leq x + \log m) \\
\approx \sum_{i=0}^{k-1} \exp(-e^{-x}) \left(\frac{e^{-x}}{i!}\right)^i + \left(\frac{m}{k}\right) \left(1 - \frac{e^{-x}}{m}\right)^{m-k} (e^{-x})^k \\
= \sum_{i=0}^{k-1} \exp(-e^{-x}) \left(\frac{e^{-x}}{i!}\right)^i + \frac{m!}{(m-k)!k!} \left(1 - \frac{e^{-x}}{m}\right)^m \left(1 - \frac{e^{-x}}{m}\right)^{-k} (e^{-x})^k.
\]

As \( m \to \infty \), we have

\[
Pr(mG_{k+1} \leq x + \log m) \to \sum_{i=0}^{k-1} \exp(-e^{-x}) \left(\frac{e^{-x}}{i!}\right)^i + \frac{1}{k!} \exp(-e^{-x}) (e^{-x})^k \\
= \sum_{i=0}^{k} \exp(-e^{-x}) \left(\frac{e^{-x}}{i!}\right)^i.
\]

Therefore we can write

\[
Pr(mG_r \leq x + \log m) \approx \sum_{i=0}^{r-1} \exp(-e^{-x}) \left(\frac{e^{-x}}{i!}\right)^i.
\]

Actually, to approximate the distribution of the Fisher statistic as we computed in the previous sections, we would let \( y = (x + \log m)/m \). Then we have

\[
Pr(G_r > y) \approx 1 - \sum_{i=0}^{r-1} \exp(-e^{-x}) \left(\frac{e^{-x}}{i!}\right)^i
\]

where \( x = my - \log m \).

Section 5: Applications

We would like to test the Fisher statistic by testing computer generated data. We will analyze seven simple models. We would like to answer the following questions:

(1) How often will we reject \( H_0 \) given \( H_0 \) is true?
(2) If we reject $H_0$, does the Fisher Test always give us the correct frequency in the data?

(3) What happens if the data does not contain Fourier frequencies?

Recall that we are testing (for the $k^{th}$ Fisher statistic)

$H_0$: There are no periods in the model; i.e., $x_t = \mu + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma^2)$.

$H_1$: There exists at least $k$ sinusoidal components in the model.

If, for a given $r$, the p-value of the test is "small," then we would reject $H_0$. By "small," we mean smaller than the given level $\alpha$. We test $G_1$ first, then $G_2$, etc., until we find that $G_r$ is no longer statistically significant. Later, we will examine a method given by Michael Shimshoni (1971).

For each of the following seven simple models, $n = 1001$ data points were generated for analysis. These models do not include mean or phase components since they would have no bearing on the analysis. Also, as before, we will denote

$$G_r = \frac{Y_r}{\sum_{j=1}^{m} Y_j}$$

where $Y_r = r^{th}$ largest of $\{|J_j|^2 : j = 1, \ldots, m\}$.

Model 1

$$x_t = \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

We computer-generated data from Model 1 one hundred times. We expect that, given level $\alpha$, we would not reject $H_0$ $100(1 - \alpha)\%$ of the time. Table 1 gives a frequency distribution of the p-values for $G_1$, for each data set.
Table 1

<table>
<thead>
<tr>
<th>p-value</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-.099</td>
<td>7</td>
</tr>
<tr>
<td>.10-.199</td>
<td>8</td>
</tr>
<tr>
<td>.20-.299</td>
<td>11</td>
</tr>
<tr>
<td>.30-.399</td>
<td>7</td>
</tr>
<tr>
<td>.40-.499</td>
<td>9</td>
</tr>
<tr>
<td>.50-.599</td>
<td>11</td>
</tr>
<tr>
<td>.60-.699</td>
<td>16</td>
</tr>
<tr>
<td>.70-.799</td>
<td>9</td>
</tr>
<tr>
<td>.80-.899</td>
<td>14</td>
</tr>
<tr>
<td>.90-.999</td>
<td>8</td>
</tr>
</tbody>
</table>

A frequency histogram (Graph 1A) of the above p-values can be found in the appendix of graphs. Let $U(G_1) = \Pr(G_1 > \hat{G}_1)$, where $\hat{G}_1$ is the observed value for $G_1$. If we find that $U(G_1)$ is a $U(0,1)$ random variable, then we can be satisfied that the probability of Type-I error is truly the set level $\alpha$ for each test. We performed a $\chi^2$-Goodness of Fit test to determine whether or not $U(G_1)$ came from a $U(0,1)$ distribution.

Let $n =$ Number of observations

Let $N_i =$ Number of p-values in interval (.1(i), .1(i+1))

Let $p_i =$ Probability of p-value in interval (.1(i), .1(i+1)), $i = 0, \ldots, 9$

Let $Q_9 = \sum_{i=0}^{9} \frac{(N_i - np_i)^2}{np_i}$

$Q_9$ has an approximate distribution of $\chi^2(9)$.

We wish to test

$H_0: U(G_1) \sim U(0,1)$

$H_1: U(G_1)$ not $U(0,1)$

We find that the observed value of $Q_9$ is 8.2 and that $\Pr(Q_9 > 8.2) \approx .49$.

Thus we do not reject $H_0$; i.e., we believe that $U(G_1) \sim U(0,1)$. 29
Graph 1B is a plot of the $|J_j|^2$ generated by the data set 55. Notice that the strongest peak is at Fourier frequency $\omega(85)$ (and at $\omega(916)$). One might hypothesize that the data had period $\frac{2\pi(85)}{1001} = .5335$ units. Before the Fisher statistic, the F statistic was used to test for periodicities. That test would have the test statistic (with the incorrect distribution) as follows:

$$F^* = \frac{|J_{85}|^2 / 2}{\left(\sum_{j=1}^{500} |J_j|^2 - |J_{85}|^2\right) / 998} \sim F(2, 998)$$

After transforming the data from data set 55, we found that $|J_{85}|^2 = 8742.73$ and $\sum_{j=1}^{500} |J_j|^2 - |J_{85}|^2 = 500669.27$. Thus $F^* = 8.715$.

Then $Pr(F^* > 8.715) = .00018$. We would strongly reject $H_0$ at level $\alpha = .05$. However, this is not adequate, for if we use the Fisher statistic, we find that $\hat{G}_1 = .01717$ and that $Pr(G_1 > .01717) = .08498$. In this case, we would not reject $H_0$ at level $\alpha = .05$. The above F-Test is too lenient. Its problem lies in the fact that we based the test statistic on the largest $|J_j|^2$ as if it were a particular $|J_j|^2$ with a pre-determined $j$ (see Section 6: cross-validation). In this case, $F^*$ defined above, does not have an F distribution.

**Model 2**

$$x_t = \varepsilon_t, \quad \varepsilon_t \sim U(-1, 1)$$

Here we expect to not reject $H_0$ by a good margin since the data should be randomly scattered. The results are in Table 2. Note: All p-values computed hereafter are accurate to five decimal places.

<table>
<thead>
<tr>
<th>Table 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed $G_r$</td>
<td>p-value</td>
<td>$\omega(j)$</td>
</tr>
<tr>
<td>$G_1 = .01329$</td>
<td>.47732</td>
<td>$j = 110$</td>
</tr>
</tbody>
</table>
Note: Using the approximate distribution of the Fisher statistic, we have
\[ \Pr(G_1 > 0.01329) \approx 0.47809. \] Hence we would not reject \( H_0 \) at level \( \alpha = 0.05 \).

In fact, we would not reject \( H_0 \) up to level \( \alpha = 0.47732 \).

Had we looked at Graph 2, a dotplot of the \( |J_j|^2 \), before this test, we would have found a strong peak at \( \omega(110) \); but, as we can see, a strong peak does not imply that the model contains sinusoidal components.

**Model 3**

\[ x_t = 0.5 \cos \left( \frac{2\pi(309)}{1001} t \right) \]

Note: \( \frac{2\pi(309)}{1001} \) is the 309th Fourier frequency when \( n = 1001 \).

Here we not only expect to reject \( H_0 \) (with 100% assurance); but, we also expect to find that the strongest frequency occurs at \( \omega(309) \). Our results are as follows:

<table>
<thead>
<tr>
<th>Observed ( G_r )</th>
<th>p-value</th>
<th>( \omega(j) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_1 = 1.0 )</td>
<td>0.0</td>
<td>( j = 309 )</td>
</tr>
</tbody>
</table>

Note: Using the approximate distribution of the Fisher statistic, we also have
\[ \Pr(G_1 > 1.0) \approx 0.0. \] Hence we get our anticipated results.

Graph 3, a dotplot of \( |J_j|^2 \), is in agreement, showing a strong peak at \( \omega(309) \).

**Model 4**

\[ x_t = 0.5 \cos \left( \frac{2\pi(309)}{1001} t \right) + \epsilon_t \]

where \( \epsilon_t \sim N(0, 1) \) for all \( t = 1, 2, \ldots, 1001 \).

We expect that our results should be similar to those with Model 1 even though we have added random \( N(0, 1) \) errors. Our results are as follows:
Table 4

<table>
<thead>
<tr>
<th>Observed $G_r$</th>
<th>p-value</th>
<th>$\omega(j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_1 = .07949$</td>
<td>0.0</td>
<td>$j = 309$</td>
</tr>
<tr>
<td>$G_2 = .01149$</td>
<td>.47369</td>
<td>$j = 470$</td>
</tr>
</tbody>
</table>

Note: Using the approximate distribution of the Fisher statistic, we have

$\Pr(G_1 > .07949) \approx 0.0$ and $\Pr(G_2 > .01149) \approx .47486$.

Hence at level $\alpha = .05$, we would reject $H_0$, as before, at frequency $\omega(309)$;

but, we would not reject $H_0$ at frequency $\omega(470)$.

Again, a dotplot of $|J_j|^2$ (Graph 4) agrees with this test showing a strong peak

at $\omega(309)$.

Model 5

$$x_t = .5 \cos \left( \frac{2\pi(617)}{1001} t \right)$$

Note: $\frac{2\pi(617)}{1001}$ is the $617^{th}$ Fourier frequency when $n = 1001$.

This time we find that there is one period; but, we do not find that the period is

$\frac{1001}{617} = 1.62$ units as expected. Because of aliasing and symmetry, we only examine

$|J_1|^2, \ldots, |J_{600}|^2$. Since we could have modeled the data as

$$x_t = -.5 \cos \left( \frac{2\pi(384)}{1001} t \right),$$

and since $|J_{617}|^2 = |J_{384}|^2$, we expect the test to conclude that the period is

$\frac{1001}{384} = 2.61$ units. The results follow:

Table 5

<table>
<thead>
<tr>
<th>Observed $G_r$</th>
<th>p-value</th>
<th>$\omega(j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_1 = 1.0$</td>
<td>0.0</td>
<td>$j = 384$</td>
</tr>
</tbody>
</table>

Note: Using the approximate distribution of the Fisher statistic, we also have

$\Pr(G_1 > 1.0) \approx 0.0$. 

32
We would reject $H_0$ at the frequency $\omega(384)$. We would thus conclude that the period is $\frac{1001}{384} = 2.61$, and not 1.62 units. This unfortunate aliasing effect is due to sampling along the integers. Notice, however, that the dotplot of $|J_j|^2$ (Graph 5) does show strong peaks at both $\omega(384)$ and $\omega(617)$.

Model 6

$$x_t = 0.2 \cos \left( \frac{2\pi(309)}{1001} t \right) + 0.4 \cos \left( \frac{2\pi(617)}{1001} t \right)$$

Under model 4, we expect that our results should show that $Y_1 = |J_{384}|^2$ (since $|J_{384}|^2 = |J_{617}|^2$) and that $Y_2 = |J_{309}|^2$. That is, we expect this since the coefficient of the sinusoidal component with Fourier frequency $\frac{2\pi(309)}{1001}$ is smaller than the coefficient of the sinusoidal component with Fourier frequency $\frac{2\pi(617)}{1001}$. Our results are as follows:

| Table 6 |
|---|---|---|
| Observed $G_r$ | p-value | $\omega(j)$ |
| $G_1 = .8$ | 0.0 | $j = 384$ |
| $G_2 = .2$ | 0.0 | $j = 309$ |

Note: Using the approximate distribution of the Fisher statistic, we have $\Pr(G_1 > .8) \approx 0.0$ and $\Pr(G_2 > .2) \approx 0.0$. Hence, at level $\alpha = .05$, we would reject $H_0$ at frequencies $\omega(384)$ and $\omega(309)$.

This time there are no surprises. We would conclude that the model has two periodic components with stronger period 2.61 units (although it is really 1.62 units) and weaker period 3.24 units. Again, Graph 6, a dotplot of the $|J_j|^2$, agrees in that it shows the strongest peak at $\omega(384)$ (and its reflection) and a somewhat weaker peak at $\omega(309)$ (and its reflection).

Model 7

$$x_t = 0.5 \cos(4t)$$
Since $\frac{2\pi(617)}{1001} \approx 4$, we expect that the results under Model 7 should be somewhat similar to those under Model 5. The results follow:

<table>
<thead>
<tr>
<th>Table 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed $G_r$</td>
</tr>
<tr>
<td>$G_1 = .80207$</td>
</tr>
<tr>
<td>$G_2 = .09478$</td>
</tr>
<tr>
<td>$G_3 = .03361$</td>
</tr>
<tr>
<td>$G_4 = .01714$</td>
</tr>
<tr>
<td>$Gr_5 = .01048$</td>
</tr>
</tbody>
</table>

We would reject $H_0$ at frequencies $\omega(362), \ldots, \omega(365)$. The real frequency is 4 which corresponds with period $\frac{2\pi}{4} \approx 1.57$. Hence, by symmetry, we have also rejected $H_0$ at frequencies $\omega(636), \ldots, \omega(639)$ which correspond with periods $p$ such that $1.57 < p < 1.59$. Thus the Fisher Test was somewhat "successful" in rejecting $H_0$ at the correct real frequency. This is, perhaps, the simplest case of a real-life situation. Rarely does data inherently have sinusoidal components with only Fourier frequencies. Therefore we will approximate the frequency (or period) as we did above. Model 7 is a good example of leakage. We find that the usual Fisher analysis gives frequencies in the vicinity of the true frequency. Graph 7 shows the results.

Section 6: Cross-Validation

While today's computers can now perform the tedious task of calculating

$$Pr(G_r > z) = m \left( \frac{m-1}{r-1} \right) \sum_{j=r}^{\infty} \left( \frac{m-r}{j-r} \right) \frac{1}{j} (-1)^{-r} (1-jz)^{m-1},$$

it may still be useful to present the topic of cross-validation (in place of the Fisher test) as a means to detect sinusoidal components in data. In cross-validation, we compute the discrete Fourier transform of only a portion (for example, 50%) of the data. From this data we would find the largest $|J_j|^2$. Since we have not yet analyzed
the remaining data, it would now be appropriate (see Section 5, Model 1) to use the F-test with the following hypotheses:

H₀: There are no periodic components in the model.
H₁: The sinusoidal component in the model has frequency ωᵣ,
    where r was pre-selected from the first portion of the data.

In essence, we cross-validate the suspected results (that ωᵣ is statistically significant) from the first portion of the data by performing statistical analysis on the second portion of the data (an F-test based on ωᵣ). For examples, we will examine the data generated by Models 1 and 4.

Model 1: We transformed the first \( n_1 = 501 \) observations. In doing so, we found that the largest and second largest \( |J_j|^2 \) values were \( |J_{15}|^2 \) and \( |J_{187}|^2 \), respectively.

We will now cross-validate the results by using an F-test on the last \( n_2 = 499 \) observations. Let \( m_2 = \frac{n_2 - 1}{2} \), as we did earlier. Like \( F_1 \) from Model 1, our test statistic, \( F_j \), will be of the following form:

\[
F_j = \frac{(|J_j|^2)/2}{\left(\sum_{i=1}^{m_2} |J_i|^2 - |J_j|^2\right)/(2m_2 - 2)} \sim F(2, 2m_2 - 2)
\]

where the \( |J_j|^2 \)'s are calculated from the last 499 observations.

In this example the data partitions are not equal (here \( n_1 = 501 \) and \( n_2 = 499 \)). Thus we must account for the change in the frequencies as follows:

\[
\omega(15) = \frac{2\pi(15)}{501} \approx \frac{2\pi(15)}{499} \text{ and } \omega(187) = \frac{2\pi(187)}{501} \approx \frac{2\pi(187)}{499}.
\]

Thus we will test \( F_{15} \) and \( F_{187} \) with \( F_j \sim F(2, 496) \).

We found that \( |J_{15}|^2 = 634.31049, |J_{187}|^2 = 1237.51709 \), and

\[
\sum_{j=1}^{499} |J_j|^2 = 132834.15625. \text{ Thus we have the following:}
\]
Table 8

<table>
<thead>
<tr>
<th>Observed $F_j$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{15} = 1.1899$</td>
<td>.30512</td>
</tr>
<tr>
<td>$F_{187} = 2.332$</td>
<td>.09816</td>
</tr>
</tbody>
</table>

Hence at level $\alpha = .05$, we do not reject $H_0$ at frequencies $\frac{2\pi(15)}{499}$ and $\frac{2\pi(187)}{499}$. This is in "agreement" with our previous results from Section 5, Model 1; that is, both tests agree that the data is not periodic.

Model 4: We transformed the data from the first $n_1 = 501$ observations. In doing so, we found that the largest and second largest $|J_j|^2$ values were $|J_{155}|^2$ and $|J_{154}|^2$, respectively.

We will now cross-validate the results by using an $F$-test on the last $n_2 = 499$ observations. Again, the data partitions are not equal. Thus to account for the change in the frequencies, we do the following:

$$
\omega(155) = \frac{2\pi(155)}{501} \approx \frac{2\pi(154)}{499} \quad \text{and} \quad \omega(154) = \frac{2\pi(154)}{501} \approx \frac{2\pi(153)}{499}.
$$

Thus we will test $F_{154}$ and $F_{153}$ with $F_j \sim F(2, 496)$.

We found that $|J_{153}|^2 = 17.10837$, $|J_{154}|^2 = 14778.26953$, and

$$
\sum_{j=1}^{249} |J_j|^2 = 13730.65625. \quad \text{Thus we have the following:}
$$

Table 9

<table>
<thead>
<tr>
<th>Observed $F_j$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{154} = 29.808$</td>
<td>0.0</td>
</tr>
<tr>
<td>$F_{153} = .30843$</td>
<td>.9697</td>
</tr>
</tbody>
</table>

Hence we strongly reject $H_0$ at frequency $\frac{2\pi(154)}{499}$ and do not reject $H_0$ at frequency $\frac{2\pi(153)}{499}$. This is in agreement with our previous results from Section 5, Model 4. In both cases we find that the period is 3.24 units.
Section 7: A Practical Guide

When Fisher originally proposed his test for periodicity, the test statistic was based upon the largest amplitude found in the periodogram. Later, Grenander and Rosenblatt extended Fisher's test for the \( r^{th} \) largest amplitude. Fisher's original alternate hypothesis, \textit{there exists at least one sinusoidal component in the model}, does not make sense when examining the \( r^{th} \) largest Fisher statistic \( G_r \), for \( r > 1 \). Logically, for example, if the first Fisher statistic \( G_1 \) is found to be significant, then when testing \( G_2 \), the alternate hypothesis should be \textit{there are at least two sinusoidal components in the model}. Thus when we test the \( k^{th} \) Fisher statistic, our alternate hypothesis is \textit{there are at least} \( k \) \textit{sinusoidal components in the model}.

Method 1

Michael Shimshoni (1971), suggests that we examine certain cut-off values. Shimshoni computed tables for the cut-offs for levels of significance \( \alpha = .01 \) and \( .05 \) for \( G_1, G_2, G_5, G_{10}, \) and \( G_{25} \). He notes that interpolation is easy for intermediate values of \( m \) and \( r \), where \( n = 2m + 1 \) is the number of observations. Shimshoni's method is as follows:

Suppose we are testing the periodicity of \( n = 2m + 1 \) observations at level \( \alpha \). Then for \( r = 1, 2, \ldots, m \), calculate the values for which \( \Pr(G_r > C_r) = \alpha \). When testing the \( r^{th} \) Fisher statistic, these values \( C_1, C_2, \ldots, C_m \) are the respective cut-off's at which we would reject \( H_0 \). Then using the data, calculate the observed values of \( G_1, \ldots, G_m \). Shimshoni would accept all frequencies corresponding to the observed \( G \) values greater than \( C_1 \). That is, he would reject \( H_0 \) at those frequencies. Say that the \( j^{th} \) amplitude was the first one less than \( C_1 \). Then Shimshoni would accept all subsequent frequencies corresponding to the observed \( G \) values found
to be greater than $C_2$. This will continue until there is an observed $G_k < C_k$. Shimshoni would not accept any other frequencies.

The overall significance level of such tests, however, is no longer $\alpha$ and there is no fixed alternate hypothesis. That is, the probability that there are at least $k$ sinusoidal components in the model is not $\alpha$. This is more of a complex decision process than a hypothesis test.

**Method 2**

Another reasonable testing method, similar to Shimshoni's, would be to sequentially test each $G_r$, for $r = 1, \ldots, m$. For the $r^{th}$ Fisher statistic, we would reject $H_0$ if the observed $G_r > C_r$, where $C_r$ is defined above. Proceed in this manner until for some $k$, there is an observed $G_k < C_k$. We would believe that there are $k-1$ sinusoidal components in the model, each with a Fourier frequency which we, respectively, found from the above testing. However, this is also a complex decision process. Note: this method and Shimshoni's method would accept the same frequencies.

**Method 3**

What if we only wish to perform one test to determine whether or not there are exactly $k$ sinusoidal components in the model? Then we would reject $H_0$ (no sinusoidal components) if $\Pr(G_1, G_2, \ldots, G_k > C_k) \leq \alpha$, where $C_k$ is defined by Shimshoni. But, since $G_1 \geq G_2 \geq \ldots, G_k$, then $\Pr(G_1, G_2, \ldots, G_k > C_k) = \Pr(G_k > C_k) = \alpha$. Thus we need only perform the $k^{th}$ test. However, this is a naive approach as it is possible for $G_i = G_{i+1}$ for some $i$. Since $C_i > C_{i+1}$, we may have the observed values $C_i > G_i = G_{i+1} > C_{i+1}$. Our previous methods would not allow us to accept the frequency associated with $G_k$, for $k = i, i+1, \ldots, m$. 
Method 4

It seems that some combination of Methods 1, 2, and 3 give the most satisfactory test procedure. A logical combination would be to reject $H_0$ and accept exactly $k$ sinusoidal components in the model if

$$\Pr(G_1 > c_1, G_2 > c_2, \ldots, G_k > c_k, G_{k+1} < c_{k+1}) = \alpha$$

for some values of $c_i$, $i = 1, \ldots, k$. However, it is clear that this is not an easy formula to solve. We could not simply use the Fisher statistic as it stands.

Therefore, since Method 1 and Method 2 give the same results, Method 3 does not always give accurate conclusions, and the statistic needed to use Method 4 has not yet been computed, we use Shimshoni’s method (or Method 2) as a practical guide.

Section 8: Additional Questions

We may thus apply the Fisher Test to models which contain periodic components and we may use cross-validation as an alternate test. This paper, however, left many questions unanswered. How close is the approximate distribution of the Fisher statistic? How quickly does the approximation converge? How well does the Fisher Test perform on periodic data that does not contain a Fourier frequency? What is the power of the Fisher Test given our alternative hypothesis? Can we formulate the Fisher statistic if the number of observations is even? These questions are not quickly solved, and at the printing of the paper, the author no longer had the time to answer those questions.
Appendix of Graphs

Graph 1A
Model 1 (Data Set 55)
Model 4

$|J_j|^2$ vs $j$
GRAPH 7

Model 7

$|J_j|^2$ vs $j$
References


VITA: Steven J. Novick

Steven Jon Novick was born at St. Barnabas Hospital in Livingston, NJ, to Marvin and Bonnie Novick. He was graduated from Scotch Plains-Fanwood High School in 1988 with honors. He received his B.S. in Mathematics and Natural Science at Muhlenberg College (5/92) in Allentown, PA. He was graduated summa cum laude from Muhlenberg College with the following honors: Phi Beta Kappa (1992), the Miriam E. Khoeler Award for Excellence in Mathematics (1991), the Morton and Mildred Sher Award for Excellence in Hebrew (1991), the Stanley D. Sloyer Award in Music (1992), and the Wesley S. Mitman Mathematical Prize (1992). He entered the graduate program at Lehigh University in Bethlehem, PA, in August 1992. Steven Novick has been studying Applied Mathematics at Lehigh University, where he has also been working as a teaching assistant. He expects to receive his M.S. in Applied Mathematics from Lehigh University in December 1994.
END OF TITLE