ILLEGIBLE DOCUMENT

THE FOLLOWING DOCUMENT(S) IS OF POOR LEGIBILITY IN THE ORIGINAL

THIS IS THE BEST COPY AVAILABLE

THE PERIODOGRAM IN HARMONIC ANALYSIS

by 500

GERALD ROBERT HOBBS JR.

B. S., California State Polytechnic College

(San Luis Obispo), 1967

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements of the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1969

Approved by:

Major Professor

LD
2668
R4
1969
H623
c.2

TABLE OF CONTENTS

1. INTRODUCTION

Time series connected with physical occurances may often appear to repeat themselves at fixed time intervals. This type of trend is referred to as cyclic or oscillatory behavior. The very nature of surrounding events may dictate some type of cycle. Thus there is little question about the applicability of harmonic analysis.

It is reasonable that we describe cyclic events with functions that are themselves cyclic. We are then led to the use of trigonometric functions and Fourier series analysis. The first section of this paper is a review of classical Fourier methods in the approximation of analytic functions by infinite sums of trigonometric terms.

Early work in the area of harmonic analysis concerned itself with the idea that cyclic behavior was often obscured by random variation within the process. Schuester was the first to devise a scheme for ferreting out these "hidden periodicities". Hotelling, Wicksell and others attacked its use on the grounds that more time series were the result of erratic shocks than were the result of cyclic factors and thus variation in the form of irregular jerks would be more prevalent than variation in the form of smooth harmonics. Schuester's idea of the periodogram persisted, however. It was used more in theory than in practice because of the great amount of computational labor involved.

The popularity of the periodogram diminished in the early 1950's only to be revived in the last part of the decade when the topic of spectral analysis became of interest. The weighted periodogram was seen to be a consistant estimate of the spectral density function and Fourier relationships were seen to exist between the new spectral functions and the sample autocovariance function, the sample auto correlation function, etc.

Thus the periodogram and Fourier analysis have again become useful. Furthermore, modern computing techniques have made its use as a research tool much more tractable.

2. FOURIER ANALYSIS: A METHODS APPROACH

The method of approximating the value of a function at a point by using a power series expansion, such as a Taylor or MacLaurin series, is well known. Power series, however, are not always practically applicable. Many times in statistical work we find trends which are in some sense oscillatory. When this is the case we usually resort to an infinite series of trigonometric functions in order that we may approximate the values of the trend function. Leonhard Euler (1707-1783) was the first to recognize the fact that an analytic function can be represented by an infinite series of sines and cosines. However, the real pioneering work in this field was done by J. B. J. Fourier. In 1822, he published "Theorie analytique de la chaluer". This publication made popular what is now known as Fourier series analysis. We will define a Fourier series expansion to be a series of the form

$$\frac{a_o}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx).$$

Not all functions possess valid Fourier expansions. Fortunately, however, expansions do exist for a wide class of functions arising in experimental work. Every function which has a valid Taylor series expansion also has a valid Fourier series expansion. Also some functions for which there exist no convergent Taylor expansions can be routinely expressed in a Fourier series. Without stating exactly under what conditions a valid Fourier expansion will exist we can state what is, for our purposes, the fundamental theorem of Fourier series analysis.

Theorem: If $f(x)$ is a single-valued function which has a derivative throughout the interval $-a \leq x \leq a$ except for a finite number of points at which it has finite discontinuities, and if $f(x + 2a) = f(x)$ then $f(t)$ can be represented by means of a Fourier series. This theorem is a slight modification of one attributed to Dirichlet. See Reddick and Miller (1955) for the a proof of this theorem.

The simplest of all Fourier series are those associated with functions whose periods are of length 2π. We shall consider this type of function first and then by means of a simple transformation generalize our results to functions with any finite period.

2.1 Derivation of the Constants

We have stated in effect that under certain conditions a function $f(x)$ can be expanded in a series of the form

$$\frac{a_o}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \cos nx).$$

Naturally then the question arises as to how to determine the coefficients a_o, a_n and b_n for all positive integral values of n.

The coefficients of a Fourier series may be calculated using the following equations

$$\text{(a)} \qquad a_n = \frac{1}{\pi} \int_c^{c+2\pi} f(x) \cos nx \, dx \text{ for } n = 0, 1, 2, \cdots$$

$$\text{(b)} \qquad b_n = \frac{1}{\pi} \int_c^{c+2\pi} f(x) \sin nx \, dx \text{ for } n = 1, 2, \cdots \qquad (1)$$

$$\frac{a_o}{2} + a_1 \cos x + a_2 \cos 2x + \cdots + a_n \cos nx$$
$$+ b_1 \sin x + b_2 \sin 2x + \cdots + b_n \sin nx \qquad (2)$$

is uniformly convergent to f(x) in the closed interval from $c \leq x \leq c + 2\pi$.

The derivation of the equations (1) is relatively straight forward as can be seen from the following.

First since we know that (2) is uniformly convergent for $c \leq x \leq c + 2\pi$ and that sin $n\pi$ and cos $n\pi$ are periodic functions we have that the series (2) is convergent for all real values of x. We can, therefore find the integral of f(x) by integrating (2) term by term.

$$\int_c^{c+2\pi} f(x)\,dx = \int_c^{c+2\pi} \frac{a_o}{2}\,dx + \int_c^{c+2\pi} a_1 \cos x dx + \cdots$$
$$+ \int_c^{c+2\pi} a_n \cos nx dx + \int_c^{c+2\pi} b_1 \sin x dx + \cdots$$
$$+ \int_c^{c+2\pi} b_n \sin nx dx.$$

We know that

$$\int_c^{c+2\pi} \sin nx dx = -\frac{1}{n} \cos nx \Big|_c^{c+2\pi} = 0 \qquad n \neq 0$$

and (3)

$$\int_c^{c+2\pi} \cos nx dx = \frac{1}{n} \sin nx \Big|_c^{c+2\pi} = 0 \qquad n \neq 0$$

All we have left then is

$$\int_c^{c+2\pi} f(x)\,dx = \int_c^{c+2\pi} a_o/2\,dx = a_o\pi$$

Thus

$$a_o = \frac{1}{\pi}\int_c^{c+2\pi} f(x)\,dx$$

Therefore equation (1a) gives us the value of a_o if we set $n = 0$.

If we let $S_n = \frac{a_o}{2} + \sum_{n=1}^{\infty} (a_j \cos jx + b_j \sin jx)$ we have that for any $\epsilon > 0$ there exists an integer $N > 0$ such that $|f(x) - S_n(x)| < \epsilon$ whenever $n \geq N$ since $S_n(x)$ converges to $f(x)$. Also, since $|\cos nx| \leq 1$ for all real values of n and x, we have that

$$|f(x) \cos nx - S_n(x) \cos nx| \leq |f(x) - S_n(x)| < \epsilon$$

whenever $n \geq N$. This means that $S_n(x) \cos nx$ converges to $f(x) \cos nx$ and thus we integrate term by term as was done previously.

$$\int_c^{c+2\pi} f(x) \cos nx dx = \int_c^{c+2\pi} \frac{a_o}{2} \cos nx + \cdots + \int_c^{c+2\pi} a_1 \cos x \cos nx dx + \cdots$$

$$+ \int_c^{c+2\pi} a_n \cos^2 nx dx + \cdots + \int_c^{c+2\pi} b_i \sin x \cos nx dx + \cdots$$

$$+ \int_c^{c+2\pi} b_n \sin nx \cos nx dx + \cdots \tag{4}$$

We can use the relations in (3) in addition to the following identities:

$$\int_c^{c+2\pi} \sin mx \cos nx dx = 0$$

$$\int_c^{c+2\pi} \cos mx \cos nx dx = 0 \quad m \neq n$$

$$\int_c^{c+2\pi} \cos^2 nx dx = \pi \quad n \neq 0$$

to reduce (4) to

$$\int_c^{c+2\pi} f(x) \cos nx dx = a_n \pi$$

which can be written as

$$a_n = \frac{1}{\pi} \int_c^{c+2\pi} f(x) \cos nx dx.$$

This completes the derivation of relation (1a).

In order to derive (1b) it is only necessary to multiply f(x) and $S_n(x)$ by sin nx and argue as before to show that

$$b_n = \frac{1}{\pi} \int_c^{c+2\pi} f(x) \sin nx dx$$

Thus the equations (1) are valid.

In the experimental case where a curve is to be fitted to a finite number of points the coefficients are determined by the same equations with the integral signs replaced by summation signs. Note that if we have N points it is only necessary to calculate N/2 terms in order to have a perfect fit. This is the Fourier analogue to the well known fact that N points can be fitted exactly by an (N - 1)th degree equation.

A Fourier series gives the value of a function, f, at a point, x, to be

$$\lim_{\epsilon \to 0} \frac{1}{2} [f(x+\epsilon)+f(x-\epsilon)]$$

Therefore, if a function, f(x) has a "jump" discontinuity at some point a, then the Fourier series representing f(x) will give a value half-way between $f(a^-)$ and $f(a^+)$. For example the Fourier representation of

$$f(x) = 0 \qquad x \leq 0$$

$$f(x) = 1 \qquad x > 0$$

will be equal to $\frac{1}{2}$ at the point x = 0.

It is of interest to note that some functions can be represented by a series of sines alone or a series of cosines alone. In other words, either all of the a_n's or all of the b_n's will vanish for some series. In general, the a_n's will vanish whenever f(x) is an odd function, i.e., f(-x) = -f(x). In this case

$$b_n = \frac{2}{\pi} \int_c^{c+\pi} f(x) \sin nx dx.$$

In a like manner, if f(x) is an even function (f(-x) = f(x)) then the b_n's will all be equal to zero and

$$a_n = \frac{2}{\pi} \int_c^{c+\pi} f(x) \cos nx dx$$

These relations are important in that their use considerably reduces the labor involved in computing the coefficients.

2.2 Change of Interval

It is obvious that the previous observations cover only a very narrow class of functions. Namely, those whose period is exactly 2π. In order to make Fourier series generally useful a way must be found to extend the range of coverable periods. Fortunately, this can be done by means of a simple transformation. Given a function $f(x)$ in an interval of length $2L$, $-L \leq x \leq L$. We can transform the function in x to a new function in z whose period is 2π. The transformation

$$z = \frac{\pi x}{L} \tag{5}$$

changes the interval to $-\pi \leq z \leq \pi$.

Now, $f(x) = f\left(\frac{Lz}{\pi}\right)$ which is function of z and can be represented by a Fourier series. That series is

$$f\left(\frac{Lz}{\pi}\right) = a_o/2 + a_1 \cos z + \cdots + a_n \cos nz + \cdots$$

$$+ b_1 \sin z + \cdots + b_n \sin nz + \cdots$$

and is valid when $-\pi \leq z \leq \pi$.

If we now apply the given transformation we have the following result:

$$f(x) = \frac{a_o}{2} + a_1 \cos \frac{\pi x}{L} + a_2 \cos \frac{2\pi x}{L} + \cdots + a_n \cos \frac{n\pi x}{L} + \cdots$$

$$+ b_1 \sin \frac{\pi x}{L} + b_2 \sin \frac{2\pi x}{L} + \cdots + b_n \sin \frac{n\pi x}{L}$$

is valid for the range $-L \leq x \leq L$.

The coefficients, the a_n's and b_n's can be computed from relations similar to those previously used

$$a_n = \frac{1}{\pi} \int_{-\infty}^{\infty} f\left(\frac{Lz}{\pi}\right) \cos nz dz = \frac{1}{L} \int_{-L}^{L} f(x) \cos \frac{n\pi x}{L} dx$$

$$b_n = \frac{1}{\pi} \int_{-\infty}^{\infty} f\left(\frac{Lz}{\pi}\right) \sin nz dz = \frac{1}{L} \int_{-L}^{L} f(x) \sin \frac{n\pi x}{L} dx \qquad (6)$$

The above may be obtained directly from (1) by substituting for z in cos nz the quantity on the right hand side of (5), using the fact that $f(x) = f\left(\frac{Lz}{\pi}\right)$, and noting the Jacobian $dz = \frac{\pi}{L} dx$.

In many cases the labor of computing the Fourier expansion can be avoided by using a combination of known series. In short if we are given a function such as

$$f(x) = x^2 + 4x + 11$$

we can find tables (e.g., C.R.C. tables pages 374-376) giving the expansions of x^2, x, and 1. Then the expansions can be multiplied by the proper constants and added to get the desired series. Thus the Fourier transform is a linear operator.

It is not necessary that the function under consideration be harmonic in order that Fourier approximations be applicable. Finite intervals of non-harmonic functions are representable in a Fourier series. Note, however, that the series thus obtained is valid only within the limits described.

For example the function

$$f(x) = x$$

can be transformed to a trigonometric series if we restrict ourselves to some finite range, say $-\pi \leq x \leq \pi$. The resulting series is (for n odd)

$$f(x) = \frac{\pi}{2} - \frac{4}{\pi}\left[\cos x + \frac{1}{3^2} \cos 3x + \frac{1}{5^2} \cos 5x + \cdots + \frac{1}{n^2} \cos nx + \cdots \right]$$

Note first of all that all of the sine terms, the b_n's, have vanished since $f(x) = x$ is an odd function. Secondly, the series converges to $f(x) = x$ only within the prescribed range. Graphically, the result of our trigonometric series would appear as in Figure 1.

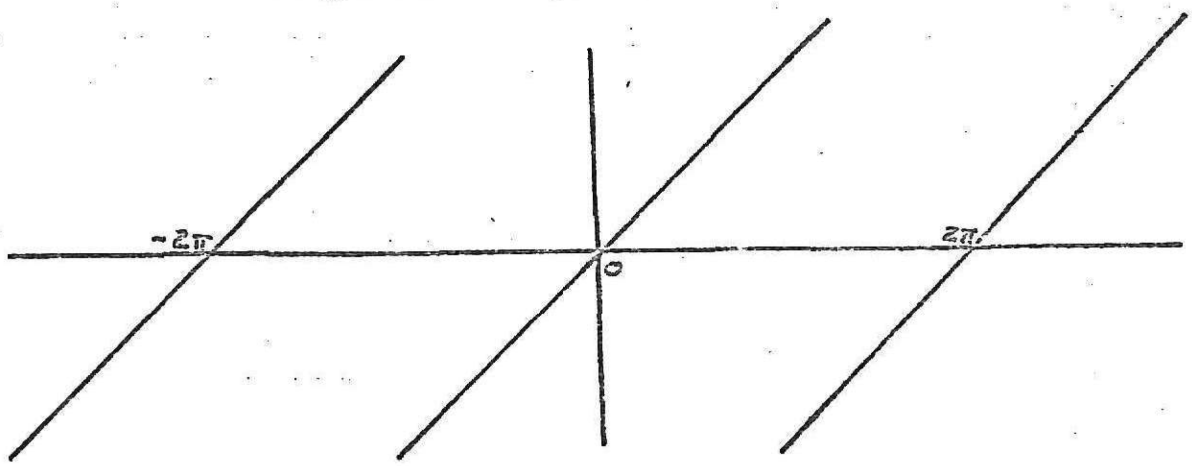

Figure 1

Obviously the resultant harmonic function is not valid out of the range $-\pi \leq x \leq \pi$.

It has been noted that the width of the interval of approximation can be changed by a linear transformation. Similarly, a transformation of the form $z = mx+k$ will transform any finite interval $a_1 \leq z \leq a_2$ to any other interval $b_1 \leq x \leq b_2$.

2.3 Transformation of a Triangular Density Function

As an example of the Fourier transformation process consider the triangular probability density function

$$f(x) = 1 + x \text{ for all } -1 \leq x \leq 0$$

$$f(x) = 1 - x \text{ for all } 0 < x \leq 1$$

$$f(x) = 0 \text{ elsewhere}$$

Assume, now, there exists a convergent Fourier sequence for the given function in the interval $-1 \leq x \leq 1$. It will be of the form

$$f(x) = \frac{a_o}{2} +, \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

In order to compute the constants, we use the equations given in (6). To find a_o we compute,

$$a_o = \int_{-1}^{0} (1 + x) \cos 0 \cdot x \, dx + \int_{0}^{1} (1 - x) \cos 0 \cdot x dx$$

$$= \int_{-1}^{0} (1 + x) \, dx) + \int_{0}^{1} (1 - x) \, dx$$

$$= \frac{1}{2} + \frac{1}{2} = 1.$$

We have then that $\frac{a_o}{2} = \frac{1}{2}$ which is called the first Fourier approximation.

Continuing, now, to find $a_1, a_2, \cdots, a_n, \cdots$

$$a_1 = \int_{-1}^{0} (1 + x) \cos \pi x dx + \int_{0}^{1} (1 - x) \cos \pi x dx$$

$$= \int_{-1}^{0} \cos \pi x dx + \int_{-1}^{0} x \cos \pi x dx + \int_{0}^{1} \cos \pi x dx - \int_{0}^{1} x \cos \pi x dx$$

$$= 4/\pi^2.$$

The general term may be written

$$a_n = \int_{-1}^{0} (1 + x) \cos n\pi x dx + \int_{0}^{1} (1 - x) \cos n\pi x dx$$

Following the above procedure we obtain

$$a_n = \frac{1}{n\pi} \sin n\pi x \Big|_{x=-1}^{0} + \frac{1}{n^2\pi^2} \cos n\pi x + n\pi \sin n\pi x \Big|_{x=-1}^{0}$$

$$+ \frac{1}{n\pi} \sin n\pi x \Big|_{x=0}^{1} - \frac{1}{n^2\pi^2} \cos n\pi x + n\pi \sin n\pi x \Big|_{x=0}^{1}$$

$$a_n = 0 + \frac{1}{n^2\pi^2} (1 - (-1)^n) + 0 + \frac{1}{n^2\pi^2} ((-1)^n - 1).$$

More succinctly,

$$a_n = \begin{cases} 4/n^2\pi^2, & \text{whenever } n \text{ is odd} \\ 0, & \text{whenever } n \text{ is even and } n \neq 0 \\ 1, & \text{when } n = 0. \end{cases}$$

Since the given function is even we expect that all of the b_n's will be equal to zero. To verify this we compute

$$b_n = \int_{-1}^{0} (1 + x) \sin n\pi x dx + \int_{0}^{1} (1 - x) \sin a\pi x dx$$

performing the desired operations yields

$$b_n = -\frac{2}{n\pi} + \frac{2}{n\pi} + \frac{(-1)^n}{n^2\pi^2} n\pi + \frac{(-1)^{n+1}}{n^2\pi^2} n\pi$$

$$= 0 \text{ for all values of } n.$$

Thus we have, for n odd, that the series

$$S_n = \frac{1}{2} + \frac{4}{\pi^2} \cos \pi x + \frac{4}{9\pi^2} \cos 3\pi x + \cdots + \frac{4}{n^2\pi^2} \cos n\pi x$$

is the Fourier transform of the given probability density function. Then, since we have selected a transformable function, S_n converges to f(x) everywhere in the range.

Graphical representation of the successive approximations to f(x) looks like:

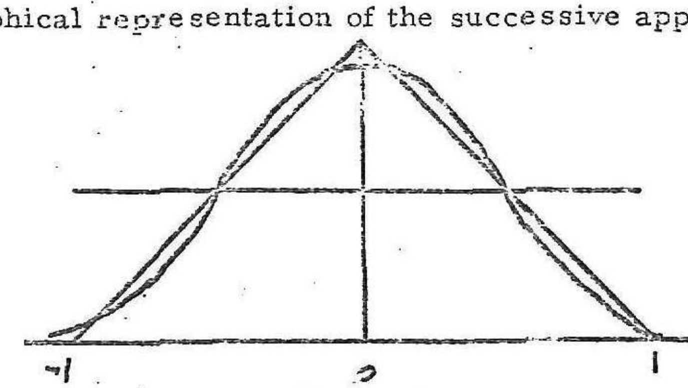

Figure 2

(a) the first approximation, $f(x) = 1/2$.

(b) the second approximation, $f(x) = \frac{1}{2} + \frac{4}{\pi^2} \cos x$.

For this particular example, the transformed Fourier series converges to f(x) very rapidly.

A by-product of Fourier analysis is the fact that we can derive some interesting arithmetical relations from them. In the example given, for instance, the value of f at zero is 1. Since S_n is uniformly convergent to f(x) everywhere in the interval $-1 \leq x \leq 1$ and

$$S_n = \frac{1}{2} + \frac{4}{\pi^2} \sum_{i=0}^{\infty} \frac{\cos((2i+1)\,nx)}{(2i+1)^2}$$

we have that

$$1 = \frac{1}{2} + \frac{4}{\pi^2}\left(1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \cdots + \cdots\right)$$

$$\frac{1}{2} = \frac{4}{\pi^2}\left(\sum_{i=0}^{\infty} \frac{1}{(2i+1)^2}\right)$$

$$\pi^2/8 = \left(1 + \frac{1}{9} + \frac{1}{25} + \cdots + \cdots\right)$$

or

$$\pi = \sqrt{8\left(1 + \frac{1}{9} + \frac{1}{25} + \cdots + \cdots\right)}$$

Thus we are provided with an expression for π which is immediately applicable to machine computation.

Fourier series analysis provides us with a useful approximation to various functions. In statistical time series analysis it becomes even more useful in that the cyclic properties of a Fourier series often correspond to harmonic variations within the series to be studied.

3. HARMONIC ANALYSIS: THE SEARCH FOR HIDDEN PERIODICITIES

Time series are usually considered to be constituted of three additive parts. They are:

a) The secular or long term movement

b) The cyclic or vibratory movement about the secular trend

c) The random variation, often called the irregular component.

The primary emphasis of this paper is with part b), the oscillatory behavior. We shall adopt Kendall's terminology with reference to the two terms "cyclic" and "oscillatory". Cyclic and its derivatives will denote the deterministic situation in which a series is strictly a function of time. In essence

$$S_t = S_{t+\omega} = S_{t+2\omega} = \cdots = \cdots$$

for some ω. Oscillatory will refer to the situation in which random variation is also present.

As has been mentioned in Section 2, most functions can be represented as the infinite sum of a series of sines and cosines. The periodic nature of trigonometric functions suggests their use in the analysis of oscillatory fluctuations. One might say that a function of time, say $x(t)$, can be represented as the sum of infinitely many harmonic terms of the form

$$A_n \cos(n\pi t/a) + B_n \sin(n\pi t/a).$$

The period of each harmonic is $2a/n$ and the amplitude is $R_n^2 = A_n^2 + B_n^2$.

Thus the periods associated with a Fourier series are

$$\frac{2a}{1}, \frac{2a}{2}, \frac{2a}{3}, \ldots, \frac{2a}{n}, \ldots$$

The early work in statistical harmonic analysis was concerned with the problem of determining the constituent periodicities of a time series. Thus the idea of Fourier series representation was intuitively appealing. Obviously the series of periods given above does not include every possible period. If we consider the Fourier sequence to be a sample from the set of all possible periods we can say that the sample is sparse in the large periods and dense for small periods. However, H. H. Turner (1913) showed that if an isolated period exists between two Fourier periods the signs of A and B will change from one period to the next. Despite the complication mentioned above Fourier series form the basis of most statistical harmonic analysis.

3.1 The Additive Property of the Intensities

One advantage of Fourier representation lies in the following relationship,

$$R_1^2 + R_2^2 + \cdots + R_n^2 = 2\sigma^2 .$$

Which can be proved, following Davis (1941), in the following manner

$$f(t) \approx \frac{1}{2} A_o + \sum_{n=1}^{N} A_n \cos \frac{n\pi t}{a} + \sum_{n=1}^{N} B_n \sin \frac{n\pi t}{a} \qquad (7)$$

That is, approximating f(t) with the first N terms of a Fourier expansion, also called the N^{th} Fourier approximation.

Let the right hand side of (7) be $f_n(t)$ and consider the integral of the square of the difference

$$I = \frac{1}{a} \int_{-a}^{a} (f(t) - f_n(t))^2 \, dt$$

$$= \frac{1}{a} \int_{-a}^{a} (f^2(t) - 2f(t)\, f_n(t) + f_n^{\,2}(t)) \, dt$$

Note that

$$\frac{1}{a} \int_{-a}^{a} \sin \frac{m\pi t}{a} \sin \frac{n\pi t}{a} \, dt = \frac{1}{a} \int_{-a}^{a} \cos \frac{m\pi t}{a} \cos \frac{n\pi t}{a} \, dt = 0, n \neq m$$

$$= 1, \; n = m.$$

We now have

$$I = \frac{1}{a} \int_{-a}^{a} f^2(t) \, dt - \left(\frac{1}{2} A_o^{\,2} + R_1^{\,2} + R_2^{\,2} + \cdots + R_n^{\,2} \right)$$

where $R_n^{\,2} = A_n^{\,2} + B_n^{\,2}$.

Since $[f(t) - f_n(t)]^2$ we have that

$$\frac{1}{2} \int_{-a}^{a} f^2(t) \, dt \geq \frac{1}{2} A_o^{\,2} + \sum_{i=1}^{N} R_i^{\,2}$$

The equality sign will hold for all functions of integrable square if $N = \infty$.

This result is known as the Bessel inequality. We then have immediately that

$$\sigma^2 = \frac{1}{2a} \int_{-a}^{a} (f^2(t) - \frac{1}{4} A_o^{\,2}) \, dt$$

$$= \frac{1}{2} \sum_{n=1}^{\infty} (A_n^{\,2} + B_n^{\,2}) = \frac{1}{2} \sum_{n=1}^{\infty} R_n^{\,2}$$

since $\frac{1}{2} A_o$ is the mean of f(t). Finally we have

$$\sum_{n=1}^{\infty} R_n^2 = 2\sigma^2$$

which is the desired result.

The graph of $y = R_n^2$ is called the periodogram. Its use in the determination of hidden periodicities was suggested by Sir Arthur Schuester in the last decade of the nineteenth century. His work was motivated by the suspicion of periodic behavior in the physical phenomena; sunspots, terrestrial magnitism, and earthquakes.

3.2 Reaction of the Intensities to a True Period

The term periodogram is misleading since y a function of frequency. The periodogram reaches its maximum value in the neighborhood of the true period. This can be illustrated by the following argument. Consider a harmonic function of the form

$$f(t) = A \sin (kt + \beta).$$

Then the Fourier approximation to f(t) yields the following values for A_n and B_n:

$$A_n = \frac{1}{a} \int_{-a}^{a} f(t) \cos \frac{n\pi t}{a} \, dt$$

$$= \frac{1}{a} \int_{-a}^{a} A \sin (kt + \beta) \cos \frac{n\pi t}{a} \, dt$$

$$= \frac{A}{a} \int_{-a}^{a} (\sin kt \cos \beta + \sin \beta \cos kt) \cos \frac{n\pi t}{a} \, dt$$

$$= \frac{A}{a} \cos\beta \int_{-a}^{a} \sin kt \cos\frac{n\pi t}{a}\, dt + \frac{A}{a} \sin\beta \int_{-a}^{a} \cos kt \cos\frac{n\pi t}{a}\, dt$$

$$= \frac{A}{2a} \cos\beta \int_{-a}^{a} (\sin(kt + \frac{n\pi t}{a}) + \sin(kt - \frac{n\pi t}{a}))\, dt$$

$$+ \frac{A}{2a} \sin\beta \int_{-a}^{a} (\cos(kt + \frac{n\pi t}{a}) + \cos(kt - \frac{n\pi t}{a}))\, dt$$

Since $\sin(kt + \frac{n\pi t}{a})$ and $\sin(kt - \frac{n\pi t}{a})$ are odd functions the first integral is equal to zero and we have that the above is equal to

$$A_n = \frac{A}{2a} \sin\beta \left(\frac{\sin(kt + \frac{n\pi t}{a})}{k + \frac{n\pi}{a}} \Bigg|_{t=-a}^{t=a} + \frac{\sin(kt - \frac{n\pi t}{a})}{k - \frac{n\pi}{a}} \Bigg|_{t=-a}^{t=a} \right)$$

$$= A \sin\beta \left(\frac{\sin(ka + n\pi)}{ka + n\pi} + \frac{\sin(ka - n\pi)}{ka - n\pi} \right)$$

Similarly we can solve for B_n

$$B_n = \frac{1}{a} \int_{-a}^{a} f(t) \sin\frac{n\pi t}{a}\, dt$$

$$= \frac{1}{a} \int_{-a}^{a} A \sin(kt + \beta) \sin\frac{n\pi t}{a}\, dt$$

$$= \frac{A}{a} \int_{-a}^{a} \sin kt \cos\beta \sin\frac{n\pi t}{a}\, dt + \int_{-a}^{a} \sin\beta \cos kt \sin\frac{n\pi t}{a}\, dt$$

$$= \frac{A}{2a} \cos\beta \int_{-a}^{a} \sin kt \sin\frac{n\pi t}{a}\, dt + \frac{A}{2a} \sin\beta \int_{-a}^{a} \cos kt \sin\frac{n\pi t}{a}\, dt$$

$$= \frac{A}{2a} \cos\beta \int_{-a}^{a} (\cos(kt - \frac{n\pi t}{a}) - \cos(kt + \frac{n\pi t}{a}))\, dt$$

$$+ \frac{A}{2a} \sin \beta \int_{-a}^{a} \sin\left(\frac{n\pi t}{a} - kt\right) - \sin\left(\frac{n\pi t}{a} + kt\right) dt$$

As before, the second integral vanishes and

$$B_n = \frac{A}{2a} \cos \beta \int_{-a}^{a} (\cos(kt - \frac{n\pi t}{a}) - \cos(kt + \frac{n\pi t}{a}))\, dt$$

$$A \cos \beta \left(\frac{\sin (ka - n\pi)}{ka - n\pi} - \frac{\sin(ka + n\pi)}{ka + n\pi}\right).$$

Thus the periodogram

$$R_n^2 = A^2 \left(\frac{\sin^2(ka + n\pi)}{(ka + n\pi)^2} + \frac{\sin^2(ka - n\pi)}{(ka - n\pi)^2}\right.$$

$$\left. - 2 \cos 2\beta \frac{\sin (ka + n\pi) \sin (ka - n\pi)}{(ka)^2 - (n\pi)^2}\right)$$

Now let the frequency $k = 2\pi/p$ and note that the n corresponding to any particular trail period, pT, is equal to $2a/T$. Using these transformations we arrive at

$$R_n^2 = R^2(T) = \frac{A^2}{4\pi^2} \left(\frac{\sin^2 2\pi(a/p + a/T)}{(a/p + a/T)^2} + \frac{\sin^2 2\pi(a/p - a/T)}{(a/p - a/T)^2}\right.$$

$$\left. - 2 \cos \beta \frac{\sin 2\pi (a/p + a/T) \sin 2\pi (a/p - a/T)}{(a/p + a/T)(a/p - a/t)}\right)$$

Let $a/p = \mu$ and $a/T = \tau$

$$R^2(T) = \frac{A^2}{4\pi^2} \left(\frac{\sin^2 2\pi(\mu + \tau)}{(\mu + \tau)^2} + \frac{\sin^2 2\pi(\mu - \tau)}{(\mu - \tau)^2}\right.$$

$$+ 2 \cos 2\beta \; \frac{\sin 2\pi(\mu + \tau) \; \sin 2\pi(\mu - \tau)}{\mu^2 - \tau^2} \Bigg)$$

The dominating term is

$$\frac{\sin^2 2\pi(\mu - \tau)}{(\mu - \tau)^2} \; .$$

Taking the derivative we find that the maximum occurs at $\tau = \mu$. Then by L'Hospital's rule the maximum value is $4\pi^2$. Further we have that the value of $R^2(T)$ is a maximum at the point $\tau = \mu$ and its value is A^2. The periodogram then will have peaks of height equal to the amplitude of the associated harmonics.

Since the sum of the R_n^2 is equal to $2\sigma^2$ we can consider the quantity

$$E_n = \frac{R_n^2}{2\sigma^2}$$

called the energy of the n^{th} harmonic, to be a measure of the relative contribution of the n^{th} harmonic to the variance of the entire system.

3.3 Construction of the Periodogram

In practice a periodogram is contructed in the following manner. First the data should be arranged in a set of equally spaced items. By previous arguments the function

$$R^2(T) = \frac{2}{N} \; A^2 + B^2 \qquad \text{(for N items)}$$

will devulge the presence of any real periods in the data. Note that the $\frac{2}{N}$ term in front of the radical is, in a certain sense, a normalizing constant.

A and B are functions of the various trial periods. That is,

$$A_n = A(T) = \sum_{t=1}^{N} X_t \sin(2\pi t/T)$$

and

$$B_n = B(T) = \sum_{t=1}^{N} X_t \cos(2\pi t/T).$$

These are the discrete analogues of the classical Fourier equations. They can be used to fit a Fourier series to a discrete set of observations. The data is then arranged in rows of length T as follows. (Called the Buys-Ballot table)

Table 1

$$\begin{array}{llll}
X_1 & X_2 & X_3 \, . \, . \, . \, . \, . \, . \, . & X_T \\
X_{T+1} & X_{T+2} & X_{T+3} \, . \, . \, . \, . \, . & X_{2T} \\
X_{2T+1} & X_{2T+2} & X_{2T+3} \, . \, . \, . \, . & X_{3T} \\
\ldots & \ldots & \ldots & \ldots \\
X_{nT+1} & X_{nT+2} & X_{nT+3} \, . \, . \, . \, . & X_{(n+1)T}
\end{array}$$

where $n+1$ is the largest multiple of T which is less than N. Sums are then taken down the columns and denoted M_1, M_2, M_3, ..., M_T. We then compute

$$A(T) = \frac{2}{N} \sum_{t=1}^{T} M_t \sin(2\pi/T) \tag{8}$$

$$B(T) = \frac{2}{N} \sum_{t=1}^{T} M_t \cos(2\pi t/T) \tag{9}$$

where $N = (n + 1)T$.

As an example, consider a set of data presented by Davis (1935), (1941) where the following analysis also appears.

Table 2

Mean Weekly Freight

Loading (1,000's)

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1919	728	687	697	715	759	809	858	892	960	967	807	758
1920	820	776	848	731	862	860	901	968	969	1005	884	723
1921	705	683	692	706	757	765	751	810	841	929	761	683
1922	702	765	826	723	787	842	825	877	935	992	944	838
1923	845	842	917	941	975	1011	986	1041	1037	1078	978	826
1924	858	908	916	875	895	906	894	974	1037	1091	975	847
1925	921	905	924	941	968	989	986	1080	1074	1107	1024	888
1926	923	919	969	958	1037	1028	1049	1104	1148	1205	1068	904
1927	946	956	1002	975	1024	999	979	1062	1097	1115	956	834
1928	862	897	951	935	1002	985	986	1058	1117	1175	1061	883
1929	893	942	962	996	1051	1052	1038	1117	1135	1169	978	835
1930	837	876	883	912	914	930	895	938	931	950	798	680
1931	719	710	735	752	740	748	738	747	737	759	655	555
1932	567	561	565	557	522	491	483	525	577	634	549	485

Table 3

Sample Calculation of R(T) for T = 15

Data Arranged as in Table 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	728	687	697	715	759	809	858	892	960	967	807	758	820	776	848
	731	862	860	901	968	969	1005	884	723	705	683	692	706	757	765
	751	810	841	929	761	683	702	765	826	723	787	842	825	877	935
	992	944	838	845	842	917	941	975	1011	986	1041	1037	1078	978	826
	858	980	916	875	895	906	894	974	1037	1091	975	847	921	905	924
	941	968	989	986	1080	1074	1107	1024	888	923	919	969	958	1037	1028
	1049	1104	1148	1205	1068	904	946	956	1002	975	1024	999	979	1062	1097
	1115	956	834	862	897	951	935	1002	985	986	1058	1117	1175	1061	883
	893	942	962	996	1051	1052	1038	1117	1135	1169	978	835	837	876	883
	912	914	930	895	938	931	950	798	680	719	710	735	752	740	748
	738	747	737	759	655	555	567	561	565	557	522	491	483	525	577
M_t =	9708	9842	9752	9968	9914	9751	9943	9948	9812	9801	9504	9322	9534	9594	9514

Note that only 165 of the 168 available numbers were used.

Using (1) and (2) we compute R(15)

$$A(15) = \frac{2}{165} \sum_{t=1}^{15} M_t \sin(2\pi t/15) = -14.1744$$

$$B(15) = \frac{2}{165} \sum_{t=1}^{15} M_t \cos(2\pi t/15) = 14.6379$$

$$R = \sqrt{A^2 + B^2} = 20.38$$

If we were to carry on with these exceedingly laborious calculations the resulting table would be

Table 4

T	R(T)	$R^2(T)$	T	R(T)	$R^2(T)$
5	5.70	32.45	16	14.82	219.50
6	46.67	2177.69	17	12.72	161.77
7	4.41	19.45	18	19.06	363.27
8	7.07	49.97	19	8.08	65.24
9	14.27	203.61	20	6.46	41.71
10	12.12	146.85	21	9.81	96.16
11	8.83	77.91	22	11.36	129.13
12	75.28	5666.44	23	17.69	312.97
13	7.87	62.00	24	18.92	358.08
14	8.84	78.19	25	2.71	7.35
15	20.38	415.18	26	6.87	47.21

If the ordered pairs (T, R(T)) were plotted and lines drawn to connect the points the resulting figure would be the periodogram

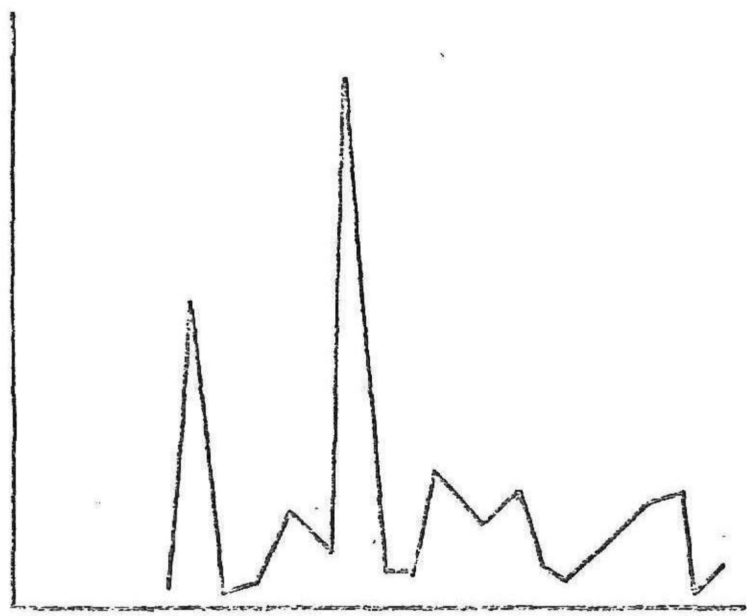

Figure 3

Thus we have completed the construction of the periodogram. Obviously the number of calculations involved make it an exceedingly unappealing device if a computer is not available.

3.4 Tests of Significance

In the example given, the periodogram certainly suggests the existance of real harmonics of 6 and 12 months. If the peaks of the periodogram had been less pronounced, however, it would be difficult to indicate which harmonics were significant and which were not. One fairly easy way is to make use of the previously mentioned concept of energy. In short, we know that the sum of the R_n^2 is equal to $2\sigma^2$ so the ratio of R_i^2 to $2\sigma^2$ would be the relative contribution of the i^{th} harmonic to the total variance of the system. In the example given above the 6 and 12 month harmonics together account for about $16\frac{1}{2}$ % of the total variance. This method leaves something to be desired at least from the classical point of view in that we still can't determine exactly which periods are important. There are, however, several methods of hypothesis testing which take the more classical approach.

3.4.1 Schuester's Test

Schuester (1898) presented a test of significance at the same time he presented his original work on the periodogram. His argument is based upon the assumption of normality in the original observations. The A(T) and B(T) are linear functions of normal observations and will thus themselves be normal. R_i^2 then is a chi-square random variable. Given this set of assumptions we define the Schuester probability that any $R^2(T)$ will exceed kR_m^2 to be equal to e^{-k}. Or

$$P(R^2(T) > kR_m^2) = e^{-k}$$

where $R_m^2 = \frac{4\sigma^2}{N}$.

To apply this test we choose the α-level we want and set it equal to e^{-k} thus determining k. Then k times our calculated value of R_m^2 serves as a lower bound for our critical region of size α.

Using the Schuester test of the preceding example we find that R_n^2's exceeding 1704 are significant at the 5% level since $e^{-3} = .05$ (approximately) and $R_m^2 = 568$. $R^2(6)$ and $R^2(12)$ are seen to be significant.

3.4.2 Walker's Test

Walker, in 1914, devised another test for the significance of periods based on the largest observed intensity.

The Walker probability that at least one intensity will exceed kR_m^2 is given by

$$P(R_{max}^2 > kR_m^2) = 1 - (1 - e^{-k})^{N/2}$$

The derivation of the above relation is based upon the Schuester criterion. The Schuester probability that any one $R^2(T)$ will exceed kR_m^2 is e^{-k}. Thus $(1 - e^{-k})$ is the probability that $R^2(T)$ is less than kR_m^2. Then $(1 - e^{-k})^{N/2}$ is the probability that all of the R_i^2's are less than kR_m^2 since a data set consisting of N points can be exactly fitted by N/2 cosine terms and $\frac{N}{2}$ sine terms. Finally $1 - (1 - e^{-k})^{N/2}$ is the probability that at least one will exceed the given limit. Tables are available for the Walker probability that at least one of the $R^2(T)$ will be greater than kR_m^2. When applied to the data given earlier, the Walker criterion for the significance of a period at the 5% level is that $R^2(T) > (7.4)(568) = 4203$, where 7.4 is the necessary table value and $R_m^2 = 568$. Only R(12) is seen as significant using this test.

3.4.3 Fisher's Test

The tests of Walker and Schuester assume not only that the observations are normally distributed but that the variance of the system is known. If the variance is unknown other methods of analysis must be found.

Fisher, in 1929, presented the statistic

$$g' = R^2 / 2S^2$$

Where R^2 is the maximum of the squared amplitudes and $2S^2 = \sum_{i=1}^{k} R_i^{\,2}$.

Then the Fisher probability that g' will be less than some given value of g is given by

$$P(g' < g) = \sum_{p=0}^{r} (-1)^p \binom{k}{p+1} (1 - (p+1)\,g)^{k-1}$$

where r is the largest integer less than $\frac{1}{g}$ and k is the largest number less than or equal to $\frac{N-1}{2}$.

The test may be readily extended to the m^{th} largest of the g' values by

$$P(g'_m < g) = \frac{k!}{(m-1)!} \sum_{p=m}^{r} \frac{(-1)^{P-M}(1-pg)^{k-1}}{p(k-p)!\ (p-m)!}$$

where the variables are defined as before.

3.4.4 Hartley's Test

Kendall (1946) made the observation, "All the tests we have described here are based on random normal variation in the original series; but in practice nobody would embark on the labour of a periodogram analysis unless he had satisfied himself that the data were not random. It seems to me that these tests are off the main point, being tests based on a hypothesis we have already rejected. They are not without their usefulness, however. We may assume with some confidence that if a particular intensity in the series is not shown as significant on the hypothesis of random variation, it is not significant when the series is systematic. What does not follow is that if one intensity is significant then others must be so even if they exceed the significance values; for they are not independent of the significant value, at least for short series. What we ought to do, perhaps, is to extract the component which is considered significant for the series and then analyse the remainder; and so on as long a significant terms appear. But this is hardly a practical computational possibility. Tests of significance in the periodogram as in the correlogram remain undiscovered".

H. O. Hartley (1949) in response to Kendall's observations presented a test for significance of the largest intensity. The test is for the maximum ratio of the quantity

$$F_{max} = \frac{1}{4} n \quad S^2_{max} (n - 2m - 1)/R^2$$

for $m \leq \frac{1}{2}(n - 1)$ and where S^2_{max} is the largest observed intensity and R^2 is the residual sum of squares.

F_{max} is distributed as an F distribution since it is the ratio of two independent chi-square random variates.

$$P(F_{max} \leq F^*) = \int_0^\infty \phi_v(s)(1 - \exp[-s^2F^*])^k \, ds$$

where $\phi_v(s)$ is the distribution of the sample standard deviation of the original observations based on v degrees of freedom. Since $v = n - 2m - 1$ the F_{max} test has 2 and $n - 2m - 1$ degrees of freedom. The Hartley or F_{max} test has been discussed at length in a previous Master's report by Backman (1963).

3.5 Interpretation of Results

The periodogram presents several practical problems to the experimenter. One of which is the tendency for false periods to be observed on either side of true peaks. Kendall's comments on lack of independence underscore this point well. In practice these secondary harmonics are usually ignored with the nagging realization that they could be true periods. A second situation which can cause difficulty arises when the periodogram takes the form indicated by Figure 3.

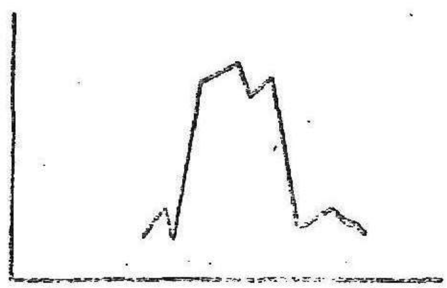

Figure 4

In Figure 3 there seems to be only one peak, yet the breadth suggests that two or more periods may be super-imposed upon each other. There are inexact methods of determining the expected width of a peak. In practice, however, certain smoothing techniques are used which increase the resolvability of the periodogram.

Also, there are sampling problems. The experimenter can never discount the existence of periods less than or equal to the sampling period or periods which are greater than or equal to one-half of the total length of the interval of observation.

Finally, there is the previously mentioned characteristic of Fourier sequences, namely that they tend to densely test for short periods but sparsely test for longer periods.

As has been seen, the periodogram tends to be very irregular in appearance. Thus it is natural to consider what might happen if the data were smoothed by means of a moving average. The resulting system has a much lower variance and thus the energy of some periods are amplified, however, spurious results are often introduced. Therefore, smoothing is not used except when added resolvability is needed.

E. T. Whittaker and G. Robinson (1940) have suggested a periodogram in which the ordinates of the Schuester periodogram are replaced by

$$n(T) = \sqrt{\sigma^2_m(T)/\sigma^2}$$

where $\sigma^2 m(T)$ is the variance of the corresponding sequence of M_i's in the Buy's-Ballot table. Results of the analysis of Whittaker-Robinson periodogram have been shown to be nearly always equal to those attained by the Schuester method.

Periodogram analysis in a certain sense the harmonic analogue of the analysis of variance. The energy or variance is attributed to cyclic effects. Thus, just as one might attribute some effect to a feed in the regular AOV we can test to see if the seasonal effect is significant over time.

4. SPECTRAL ANALYSIS

The use of the periodogram as a research tool has been largely forgotten. Few journal articles have been published on it since the early nineteen-fifties. It was impractical during its popularity because of the inordinate number of calculations that need to be performed. By the time computing devices had become highly sophisticated interest in the periodogram had waned. In a recent journal article Jones (1965) indicates that the periodogram approach to time series analysis is, with modern computes, often the most efficient way to carry out a spectral or harmonic analysis. It would thus not be suprising to see it reappear as an applied tool.

The present importance of the periodogram lies in the fact that it forms the basis for spectral analysis. The old idea of "the scheme of hidden periodicities" differs little from "An expression of the second moments of an ensemble process in terms of frequencies" as spectral analysis is defined to be by Blackman and Tukey (1958). Also, the concept of energy defined earlier in this paper corresponds to the modern idea of power.

4.1 Lack of Consistancy

Note that the periodogram can be expressed in the following form

$$I_n(\lambda) = \frac{2}{N} \left| \sum_{t=1}^{N} X_t e^{it\lambda} \right|^2$$

using the Euler relationship $\cos x + i \sin x = e^{ix}$. Then we can show, following Hannan (1960) that the expectation

$$E(I_n(\lambda)) = 4\pi f(\lambda) \qquad \text{(almost everywhere)}$$

To see this define the power spectrum $f(\lambda)$ as

$$\lim_{n \to \infty} \frac{1}{4\pi} I_n(\lambda) = f(\lambda) \tag{10}$$

and note that

$$E[(\frac{1}{4\pi}) I_n(\lambda)] = \frac{1}{2n\pi} E\Big[\sum_{t=1}^{N} \sum_{s=1}^{N} x(s)\, x(t)\, e^{i(t-s)\lambda} \Big]$$

Assuming that the $x(t)$ is stationary in the wide sense (covariance stationary) we have

$$E[\frac{1}{4\pi} I_n(\lambda)] = \frac{1}{2\pi} \sum_{t=1}^{N} \sum_{s=1}^{N} \frac{\gamma(t-s)}{n} e^{i(t-s)\lambda} \tag{11}$$

The double summation in (11) contains n^2 terms of which only $2n - 1$ are distinct. Therefore we may rewrite it as a single summation with the proper counting factor. That is

$$E[\frac{1}{4\pi} I_n(\lambda)] = \frac{1}{2\pi} \sum_{t=1-n}^{n-1} (1 - \frac{|t|}{n})\, \gamma(t)\, e^{it\lambda}$$

which is the n^{th} Cesaro mean for $f(\lambda)$.

Furthermore, we see from equation (11) that the periodogram is the Fourier transform of the sample autocovariance function.

Thus we have the desired result. This suggests the use of the periodogram as an estimator of $f(\lambda)$. However, Hannan further shows that

$$Var\{I_n(\lambda)\} = 4\sigma^4 + \frac{4k_4}{N} + \frac{4\sigma^4 \sin^2 N\lambda}{N^2 \sin^2 \lambda}$$

where k_4 is the fourth cumulant of $\{x_t\}$. And

$$\mathrm{Cov}\{I_n(\lambda_1), I_n(\lambda_2)\} = \frac{4k_4}{N} + \frac{4\sigma^4}{N^2} + \left(\frac{\sin\frac{1}{2}N(\lambda_1+\lambda_2)}{\sin\frac{1}{2}(\lambda_1+\lambda_2)}\right) + \left(\frac{\sin 1/2N(\lambda_1+\lambda_2)^2}{\sin 1/2(\lambda_1-\lambda_2)^2}\right)$$

From this result we see that $I_n(\lambda)$ has variance of order one and the covariance of $I_n(\lambda_1)$ and $I_n(\lambda_2)$ is of order N^{-1}. Thus we have that the periodogram is not a consistant estimate of $f(\lambda)$. That is, its variance does not vanish as N approaches infinity. This result explains the fact that the periodogram does not become more stable as the amount of data increases. However, the covariance approaches 0 in the limit thus we have that $I_n(\lambda_1)$ and $I_n(\lambda_2)$ are nearly independent for large N. This lack of consistancy in the periodogram has led to other methods of estimating the spectrum of a process.

4.2 Spectral Windows

P. J. Daniell (1946) suggested that the periodogram be averaged over neighboring frequencies in order to stabilize the estimate. This led to the concept of the smoothed or integrated periodogram. Wiener (1930) showed that by averaging over S frequencies and letting $S \to \infty$ as $N \to \infty$ but at a slower rate so that $\frac{S}{N} \to 0$ a consistant estimate could be found. The new estimate of $f(\lambda)$ can be written in the form

$$\int_{\lambda_1}^{\lambda_2} \omega(\lambda)\, I_n(\lambda)\, d\lambda$$

where $\omega(\lambda)$ is the weighting function used in smoothing the periodogram. Usually $\omega(\lambda)$ will be a lag function, that is, a function of $(t - \lambda)$. Recently the Fourier transform $\omega(\lambda)$ has become popularized as the spectral window.

Daniell's name has been associated with the most simple case, that is when the averaging is done with equal weights. Daniell's window is also called the rectangular window. Other weighting functions have been suggested by Parzen, Blackman and Tukey, et al. Some examples can be seen in Figure 3.

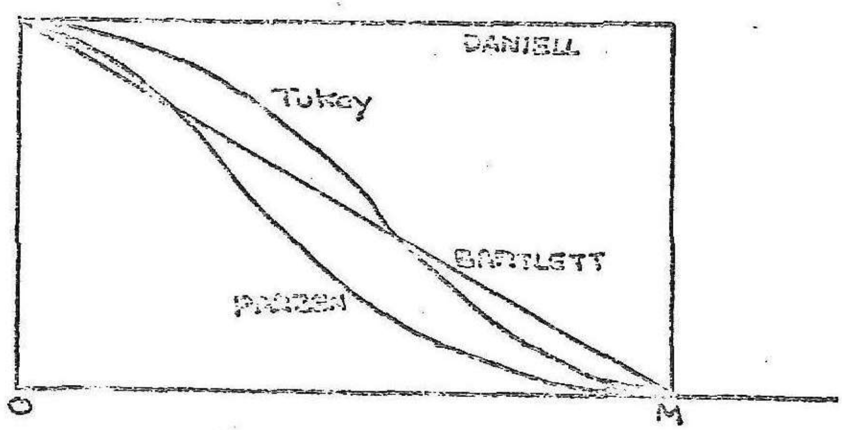

Figure 5

Weights:

Daniell $$w = \begin{cases} 1 & |u| \le M \\ 0 & |u| > M \end{cases}$$

Bartlett $$w = \begin{cases} 1 - \dfrac{|u|}{M} & |u| \le M \\ 0 & |u| > M \end{cases}$$

Tukey $$w = \begin{cases} \dfrac{1}{2}\left(1 + \cos \dfrac{\pi u}{M}\right) & |u| \le M \\ 0 & |u| > M \end{cases}$$

Parzen $$w = \begin{cases} 1 - 6\left(\dfrac{u}{M}\right)^2 + y\left(\dfrac{|u|}{M}\right)^3 & |u| \le \dfrac{M}{2} \\ 2\left(1 - \dfrac{|u|}{M}\right)^3 & \dfrac{M}{2} \le |u| \le M \\ 0 & |u| > M \end{cases}$$

Thus the periodogram while is itself not a consistant estimator of the spectral density function, is the basic idea from which spectral estimates were developed and thereby retains its value to statisticians.

5. CONCLUSION

Historically the periodogram traces its origin to the late 1890's and the work of Sir Arthur Schuester. His concern was to determine the periodic structure of certain physical phenomena. Hotelling and others raised objections to the assumptions underlying its use and thus the periodogram has been one of the more controversial subjects in statistical research.

As a practical tool it was unwieldly because of the numerous calculations involved. With the advent of modern computing techniques, however, the periodogram may again become useful in applied statistics. Its analysis is similar to the analysis of variance since the ideas of power or energy is related to the concept of variance. This can be extended to the idea of regression in that models can be composed using the significant periods.

One difficulty faced by experimenters is in testing for the significance of periods. None of the available tests is acceptable to everyone. Much work still needs to be done in the field of hypothesis testing.

The periodogram is most important at the present time because of its relation to the analysis of the power spectrum. It is not a consistant estimate of the spectral density function but some weighted averages of it are. The weighting functions are called spectral windows. The various shapes of the windows give different characteristics to the estimaters. Some create a higher resolvability while others yield more stability.

REFERENCES

Backman G., "Analysis of Stationary Time Series", unpublished Master's report Kansas State University (1963).

Bartlett M. S., Stochastic Processes, Cambridge Press. (1955).

Bartlett and Medki, "On the Efficiency of Proceedures for Smoothing Periodograms from Time Series with Continuous Spectra" Biometrika Vol. 42 (1955), pp. 143-150.

Blackman and Tukey, The Measurement of Power Spectra, Dover (1958).

Bochner and Kawata, "A Limit Theorem for the Periodogram", Annals of Math Stat. Vol. 29 (1958), pp. 1199-1220.

Davis, H. T., The Analysis of Economic Time Series, Principia Press (1946).

Davis & Nelson, Elements of Statistics, Principia Press (1941).

Fisher R. A., "Tests of Significance in Harmonic Analysis" Proceedings of the Royal Society of London, Series A. Vol. 125: (1929), pp. 54-59.

Greenstein B., "Periodogram Analysis with Spectral Application to Business Failures in the U. S. 1867-1932", Econometrica Vol. 3 (1935).

Grenander & Rosenblatt, Statistical Analysis of Stationary Time Series, Wiley (1957).

__________, "On Spectral Analysis of Stationary Time Series", Proceedings of the National Academy of Sciences, Vol. 38 (1952)

__________, "Statistical Spectral Analysis Arising From Stationary Stochastic Processes" Annals of Math Stat. Vol. 24 (1953) pp. 537-558.

Hannan E. J., Time Series Analysis, Methuen (1960).

__________, Notes on Time Series Analysis Part I, Department of Statistics John Hopkins University (1965).

__________, Notes on Times Series Analysis Part II, Department of Statistics, John Hopkins University (1965).

Hartley H. O., "Tests of Significance in Harmonic Analysis", Biometrika Vol. 36 (1949) pp. 194-201.

Jenkins & Watts, Spectral Analysis and its Applications, Holden-Day (1968).

Jones R. H., "A Reappraisal of the Periodogram in Spectral Analysis", Technumetrics, Vol. 7 (1965) pp. 531-542.

Kawata T., "Some Convergence Theorem for Stationary Stochastic Processes" Annals of Math. Stats., Vol. 30 (1959) pp. 1192-1214.

Moran P. A. P., "Some Theorem on Time Series I" Biometrika, Vol. 34 (1947).

Parzen, E., "On Choosing an Estimate of the Spectral Density Function of a Stationary Time Series" Annals of Math. Stat. Vol. 28 (1957) pp. 921-932.

____________, "On Consistant Estimates of the Spectrum of a Stationary Time Series" Annals of Math. Stat. Vol. 28 (1957) pp. 329-348.

Reddick & Miller, Advanced Mathematics for Engineers, Wiley (1955).

Schuester A. "On the Investigation of Hidden Periodicities with Application to a suppose 26-day period of Meterological Phenomena" Terrestrial Magnetism, Vol. 3 (1898) pp. 13-41.

Spencer-Smith J. L., "The Oscillatory Properties of the Moving Average" Journal of the Royal Statistical Society (Supp). Vol. 9 (1947) pp. 104-113.

Turner H. H., "On the Expression of Sunspon activity as a Fourier Sequence" Monthly Notices of the Royal Astronomical Society, Vol. 73 (1913) pp. 715-717.

Whittaker & Robinson, The Calculus of Observation, Blackie & Sons (1940).

Wiener N., "Generalized Harmonic Analysis" Acta Mathematica, Vol. 55 (1930) pp. 117-132.

ACKNOWLEDGEMENT

I would like to thank Dr. R. A. Waller and Dr. H. C. Fryer for their kind attention for the last two years.

THE PERIODOGRAM IN HARMONIC ANALYSIS

by

GERALD ROBERT HOBBS JR.

B.S., California State Polytechnic College

(San Luis Obispo), 1967

AN ABSTRACT OF A REPORT

submitted in partial fulfillment of the

requirements of the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1969

Abstract

Many physical situations can be considered to be developing in time. The usual idea of regression is one of polynomial relationships, in our case, with respect to time. The idea behind harmonic analysis and the periodogram is that the relationship may be more adequately described in cyclic terms. Scheuster in 1898 opened the field by applying Fourier transforms to data from the physical sciences. He presented his own test of significance as did others in the ensueing years, none of which is completely acceptable. Thus the computational difficulties coupled with the lack of really good tests of significance combined choke off interest in the field. In the late 1950's Blackman and Tukey popularized the field of spectral analysis in which the periodogram was a basic element. Weighted averages of the periodogram were known to give consistant estimates of the spectral density function and the older concept of the energy associated with a frequency was analogous to the newer idea of the power spectrum.

Interest has further been revived in the periodogram by the advent of modern computing machines. The objective of this paper is to present the old concept of the periodogram as a tool for discovering "hidden periodicities" and then relate it to spectral analysis.