THE DISTRIBUTION OF THE MULTIPLE CORRELATION COEFFICIENT IN PERIODOGRAM ANALYSIS

BY D. M. STARKEY

1. **Geometrical interpretation of the problem.** We begin with a summary of some recent work by Hotelling, in a form relevant to this particular problem.[1] He suggests that the general question of finding the distribution of the multiple correlation coefficient corresponding to a fitted regression of y upon x may be solved by evaluating definite integrals corresponding to invariants of certain curves, surfaces, etc. For the purposes of illustration we may consider the case of fitting the relation

$$Y = a + bf(x, k, \epsilon)$$

where f is an arbitrary function, and a, b, k, ϵ are constants, to the observations y, where we are given n values of $y, y_1, y_2, \cdots, y_n$ and the corresponding values of $x, x_1, \cdots, x_n$. We shall postulate that the y's are independent and normally distributed about a certain mean and that the regression may be fitted by means of the principle of least squares.

We must minimize the sum of squares

$$\sum_{\alpha=1}^{\alpha=n} (y_\alpha - Y_\alpha)^2 = \sum_{\alpha=1}^{\alpha=n} [y_\alpha - a - bf(x_\alpha, k, \epsilon)]^2$$

and hence we differentiate with respect to a, obtaining the first condition for a minimum

$$\sum_{\alpha=1}^{\alpha=n} [y_\alpha - a - bf(x_\alpha, k, \epsilon)] = 0.$$

In the following, all summations take place over a range $\alpha = 1$ to n. Then we have

$$a = \bar{y} - b\bar{f}$$

where

$$\bar{y} = \frac{\Sigma y_\alpha}{n}, \qquad \bar{f} = \frac{\Sigma f(x_\alpha, k, \epsilon)}{n}$$

Thus we minimize the sum of squares

$$\Sigma[(y_\alpha - \bar{y}) - b(f(x_\alpha, k, \epsilon) - \bar{f})]^2$$

[1] Harold Hotelling, "Tubes and spheres in n-spaces, and a class of statistical problem", *American Journal of Mathematics*, April, 1939.

or, putting $y'_\alpha = y_\alpha - \bar{y}$

$$Y'_\alpha = Y_\alpha - \bar{Y} = bf(x_\alpha, k, \epsilon) - \bar{f}$$

we see that the quantity $\Sigma(y'_\alpha - Y'_\alpha)^2$ is to be minimized.

Geometrically we may regard the set of values $(y_1, \cdots, y_n)$ as defining a point in n-space, and $(Y_1, \cdots, Y_n)$ will also represent a point in n-space on the 4-dimensional surface which may be obtained by eliminating a, b, k, ϵ from the relations $Y = a + bf(x, k, \epsilon)$. The points $(y'_1, \cdots, y'_n)$ and $(Y'_1, \cdots, Y'_n)$ represent the orthogonal projections of $(y_1, \cdots, y_n)$ and $(Y_1, \cdots, Y_n)$ on the plane $\Sigma\, y_\alpha = 0$. Hence we have to minimize the distance between these projections, noticing that $(Y'_1, \cdots, Y'_n)$ now lies on the 3-dimensional projection of the surface on which $(Y_1, \cdots, Y_n)$ lies. The multiple correlation between the observed and fitted values is defined as

$$R = \frac{\Sigma(y_\alpha - \bar{y})(Y_\alpha - \bar{Y})}{\sqrt{\Sigma(y_\alpha - \bar{y})^2 \Sigma(Y_\alpha - \bar{Y})^2}} = \frac{\Sigma y'_\alpha Y'_\alpha}{\sqrt{\Sigma y'^2_\alpha \Sigma Y'^2_\alpha}}$$

and this is equal to $\cos\theta$, where θ is the angle between the lines joining the origin to the points $(y'_1, \cdots, y'_n)$ and $(Y'_1, \cdots, Y'_n)$. For the purpose of evaluating R we may thus consider the projections of these points on the unit sphere in $\Sigma\, y_\alpha = 0$ with centre the origin, these being

$$\left(\frac{y'_1}{\sqrt{\Sigma y'^2_\alpha}}, \cdots, \frac{y'_n}{\sqrt{\Sigma y'^2_\alpha}}\right) \quad \text{and} \quad \left(\frac{Y'_1}{\sqrt{\Sigma Y'^2_\alpha}}, \cdots, \frac{Y'_n}{\sqrt{\Sigma Y'^2_\alpha}}\right),$$

As by hypothesis the distribution of y has spherical symmetry about some point on the line $y_1 = y_2 = \cdots\cdots = y_n$, then the distribution of y' has spherical symmetry about the origin, and the probability distribution of the projection of y' on the unit sphere is uniform. The projection of Y' lies on a 2-dimensional surface on the $(n - 2)$-dimensional sphere, and for a given Y' the probability that R is as great or greater than $\cos\theta$ is proportional to the volume of the sphere in the $(n - 2)$-dimensional spherical space with centre Y' and geodesic radius θ, so that the total probability that R lies between $\cos\theta$ and 1 is equal to the ratio of the "area" of the portion of the unit sphere included by the envelope of these geodesic spheres to the "area" of the unit sphere. This envelope is that part of the unit sphere in $\Sigma\, y_\alpha = 0$ which is at a geodesic distance θ from the 2-dimensional surface on which the projection of Y' lies, termed a "tube" by Hotelling.

For very small values of θ it may be assumed that this ratio is equal to the area of the two-dimensional surface on which Y' lies, multiplied by a fixed multiple of θ^{n-4}. This is fairly evident intuitively, but has recently been substantiated by some results of Weyl[2] who shows that this is correct for small values of θ, and indicates a series from which could be derived a series of ascend-

[2] H. Weyl, "On the volume of tubes," *American Journal of Mathematics*, April, 1939.

ing powers of θ^2 by which successive adjustments could be made for larger values of θ. The coefficients in this series are finite invariants of the surface in which we are working. If we accept the first approximation we must consider the question of the extent of the surface, which depends on the range of values of the parameters k, ϵ. The range which is eventually chosen depends on the needs of the practical statistician, while keeping in view the mathematical possibilities of effecting a solution. In the following work we consider in particular the case of periodogram analysis by putting $f(x, k, \epsilon) = \cos(kx + \epsilon)$.

2. **The case of periodogram analysis.** With the notation of the preceding paragraph, we fit

$$Y_\alpha = a + b\cos(kx_\alpha + \epsilon)$$

to data (x_α, y_α) $\alpha = 1, 2, \cdots, n$.
We shall assume that the variate x is a measurement of time or some other quantity for which the measurements are made at equal intervals, taken as unity for convenience, so that

$$x_1 = 0, \qquad x_2 = 1, \cdots, x_n = n - 1.$$

Now we shall see later that we are interested in values of k such that $0 < k < 2\pi$. For this range

$$\bar{f} = \frac{\Sigma\cos(kx_\alpha + \epsilon)}{n}$$

$$= \frac{\sin(\frac{1}{2}nk)\cos[\epsilon + \frac{1}{2}k(n-1)]}{n\sin(\frac{1}{2}k)}.$$

Hence, if Y'' represents the projection of Y' on the unit sphere

$$Y''_\alpha = \lambda\left[\cos(kx_\alpha + \epsilon) - \frac{\sin(\frac{1}{2}nk)\cos[\epsilon + \frac{1}{2}k(n-1)]}{n\sin(\frac{1}{2}k)}\right]$$

where λ is to be determined so that

$$\Sigma Y''^2_\alpha = 1.$$

Now

$$\Sigma Y''^2_\alpha = \lambda^2\left[\Sigma\cos^2(kx_\alpha + \epsilon) - \frac{\sin^2(\frac{1}{2}nk)\cos^2[\epsilon + \frac{1}{2}k(n-1)]}{n\sin^2(\frac{1}{2}k)}\right]$$

and

$$\Sigma\cos^2(kx_\alpha + \epsilon) = \frac{1}{2}n + \frac{1}{2}\frac{\sin nk\cos[2\epsilon + k(n-1)]}{\sin k}$$

and hence

$$\lambda = \frac{1}{\sqrt{\frac{1}{2}n + \frac{1}{2}\frac{\sin nk\cos[2\epsilon + k(n-1)]}{\sin k} - \frac{\sin^2(\frac{1}{2}nk)\cos^2[\epsilon + \frac{1}{2}k(n-1)]}{n\sin^2(\frac{1}{2}k)}}}$$

the expression being continous at $k = \pi$.

Then $$Y''_\alpha = \lambda(\cos(kx_\alpha + \epsilon) - \bar{f})$$
$$= \lambda \cos(kx_\alpha + \epsilon) + \xi \text{ say.}$$

Regarding k and ϵ as curvilinear coordinates of a point on the surface, we apply the formula

$$\sqrt{EG - F^2}\, dk\, d\epsilon$$

for the element of surface area, where

$$E = \Sigma\left(\frac{\partial Y''_\alpha}{\partial k}\right)^2, \qquad F = \Sigma \frac{\partial Y''_\alpha}{\partial k}\frac{\partial Y''_\alpha}{\partial \epsilon}, \qquad G = \Sigma\left(\frac{\partial Y''_\alpha}{\partial \epsilon}\right)^2.$$

In evaluating these summations, we shall need the fòllowing results: $\Sigma Y''_\alpha = 0$, $\Sigma Y''^2_\alpha = 1$, from which we obtain

(1) $$\Sigma \cos(kx_\alpha + \epsilon) = \frac{-n\xi}{\lambda}$$

(2) $$\Sigma \cos^2(kx_\alpha + \epsilon) = \frac{1 + n\xi^2}{\lambda^2}.$$

Differentiating these relations, we have

(3) $$\Sigma x_\alpha \sin(kx_\alpha + \epsilon) = \frac{\partial}{\partial k}\left(\frac{n\xi}{\lambda}\right)$$
$$= \frac{n\xi_k}{\lambda} - \frac{n\xi\lambda_k}{\lambda^2}$$

(4) $$\Sigma \sin(kx_\alpha + \epsilon) = \frac{\partial}{\partial \epsilon}\left(\frac{n\xi}{\lambda}\right)$$
$$= \frac{n\xi_\epsilon}{\lambda} - \frac{n\xi\lambda_\epsilon}{\lambda^2}$$

(5) $$\Sigma x_\alpha^2 \sin^2(kx_\alpha + \epsilon) = \frac{1}{2}\Sigma x_\alpha^2 + \frac{1}{4}\frac{\partial^2}{\partial k^2}\left(\frac{1 + n\xi^2}{\lambda^2}\right)$$
$$= \frac{n(n-1)(2n-1)}{12} + \frac{1}{4}\left[\left(-\frac{2\lambda_{kk}}{\lambda^3} + \frac{6\lambda_k^2}{\lambda^4}\right)(n\xi^2 + 1) - \frac{8\lambda_k n\xi\xi_k}{\lambda^3} + \frac{2n}{\lambda^2}(\xi\xi_{kk} + \xi_k^2)\right]$$

(6) $$\Sigma x_\alpha \cos(kx_\alpha + \epsilon)\sin(kx_\alpha + \epsilon) = -\frac{1}{2}\frac{\partial}{\partial k}\left(\frac{1 + n\xi^2}{\lambda^2}\right)$$
$$= \frac{\lambda_k(1 + n\xi^2)}{\lambda^3} - \frac{n\xi\xi_k}{\lambda^2}$$

(7) $$\Sigma \cos (kx_\alpha + \epsilon) \sin (kx_\alpha + \epsilon) = -\frac{1}{2}\frac{\partial}{\partial \epsilon}\left(\frac{1 + n\xi^2}{\lambda^2}\right)$$
$$= \frac{\lambda_\epsilon}{\lambda^3}(1 + n\xi^2) - \frac{n\xi\xi_\epsilon}{\lambda^2}$$

(8) $$\Sigma x_\alpha \sin^2 (kx_\alpha + \epsilon) = \frac{1}{2}\Sigma x_\alpha + \frac{1}{4}\frac{\partial^2}{\partial k \partial \epsilon}\left(\frac{1 + n\xi^2}{\lambda^2}\right)$$
$$= \frac{n(n-1)}{4} - \frac{1}{2}\frac{\lambda_{\epsilon k}}{\lambda^3}(1 + n\xi^2) + \frac{3}{2}\frac{\lambda_\epsilon \lambda_k}{\lambda^4}(1 + n\xi^2) - \frac{n\lambda_\epsilon \xi\xi_k}{\lambda^3}$$
$$+ \frac{n\xi_k \xi_\epsilon}{2\lambda^2} + \frac{n\xi\xi_{\epsilon k}}{2\lambda^2} - \frac{n\xi\xi_\epsilon \lambda_k}{\lambda^3}$$

Now

$$\frac{\partial Y''_\alpha}{\partial k} = \lambda_k \cos (kx_\alpha + \epsilon) - \lambda x_\alpha \sin (kx_\alpha + \epsilon) + \xi_k$$

and

$$\frac{\partial Y''_\alpha}{\partial \epsilon} = \lambda_\epsilon \cos (kx_\alpha + \epsilon) - \lambda \sin (kx_\alpha + \epsilon) + \xi_\epsilon$$

so that with the above definitions of E, F, G we obtain

$$E = \lambda_k^2 \Sigma \cos^2(kx_\alpha + \epsilon) + \lambda^2 \Sigma \sin^2(kx_\alpha + \epsilon) + n\xi_k^2 - 2\lambda\lambda_k \Sigma x_\alpha \cos (kx_\alpha + \epsilon)$$
$$\cdot \sin (kx_\alpha + \epsilon) - 2\lambda\xi_k \Sigma x_\alpha \sin (kx_\alpha + \epsilon) + 2\lambda_k \xi_k \Sigma \cos (kx_\alpha + \epsilon)$$
$$= -\frac{\lambda_{kk}}{2\lambda} + \frac{\lambda_k^2}{2\lambda^2} + \frac{n(n-1)(2n-1)}{12}\lambda^2 - \tfrac{1}{2}n\lambda^2 \bar{f}_k^2 + \tfrac{1}{2}n\bar{f}^2\lambda^2 \bar{f}_{kk}$$

$$F = \lambda_\epsilon \lambda_k \Sigma \cos^2 (kx_\alpha + \epsilon) + \lambda^2 \Sigma x_\alpha \sin^2 (kx_\alpha + \epsilon) + n\xi_\epsilon \xi_k$$
$$-\lambda\lambda_k \Sigma \sin (kx_\alpha + \epsilon) \cos (kx_\alpha + \epsilon) + \xi_\epsilon \lambda_k \Sigma \cos (kx_\alpha + \epsilon) - \lambda\xi_\epsilon \Sigma x_\alpha \sin(kx_\alpha + \epsilon)$$
$$+ \lambda_\epsilon \xi_k \Sigma \cos(kx_\alpha + \epsilon) - \lambda\lambda_\epsilon \Sigma x_\alpha \sin (kx_\alpha + \epsilon) \cos (kx_\alpha + \epsilon) - \lambda\xi_k \Sigma \sin (kx_\alpha + \epsilon)$$
$$= \lambda_\epsilon \lambda_k \left(\frac{1}{2\lambda^2}\right) - \frac{\lambda_{k\epsilon}}{2\lambda} + \frac{\lambda^2 n(n-1)}{4} + \frac{n\bar{f}\bar{f}_{k\epsilon}}{2}\lambda^2 - \frac{n\lambda^2 \bar{f}_k \bar{f}_\epsilon}{2}$$

$$G = \lambda_\epsilon^2 \Sigma \cos^2 (kx_\alpha + \epsilon) + \lambda^2 \Sigma \sin^2 (kx_\alpha + \epsilon) + n\xi_\epsilon^2 - 2\lambda\xi_\epsilon \Sigma \sin (kx_\alpha + \epsilon)$$
$$- 2\lambda\lambda_\epsilon \Sigma \cos (kx_\alpha + \epsilon) \sin (kx_\alpha + \epsilon) + 2\lambda_\epsilon \xi_\epsilon \Sigma \cos (kx_\alpha + \epsilon)$$
$$= -\frac{\lambda_\epsilon^2}{\lambda^2} + n\lambda^2 - (n\bar{f}^2\lambda^2 + 1) - n\lambda^2 \bar{f}_\epsilon^2,$$

after using the relation $\xi = -\bar{f}\lambda$ to eliminate ξ.

These relations give

$$
\begin{aligned}
EG - F^2 = &\left(-\frac{\lambda_{kk}}{2\lambda} + \frac{\lambda_k^2}{2\lambda^2} + \frac{n-1}{12}\, n(2n-1)\lambda^2 - \frac{1}{2}\, n\lambda^2 \bar{f}_k^2 + \frac{1}{2}\, n\bar{f}^2\lambda^2 \bar{f}_{kk}\right) \\
&\times\left(-\frac{\lambda_\epsilon^2}{\lambda^2} + n\lambda^2 - (n\bar{f}^2\lambda^2 + 1) - n\lambda^2\bar{f}_\epsilon^2\right) \qquad (9)\\
&-\left(\frac{\lambda_\epsilon\lambda_k}{2\lambda^2} - \frac{\lambda_{k\epsilon}}{2\lambda} + \frac{\lambda^2 n(n-1)}{4} + \frac{n\bar{f}\bar{f}_{k\epsilon}\lambda^2}{2} - \frac{n\lambda^2\bar{f}_k\bar{f}_\epsilon}{2}\right)^2.
\end{aligned}
$$

The area of the surface on which Y'' lies is

$$\sqrt{EG - F^2}\, dk\, d\epsilon$$

over an appropriate range of values of k and ϵ, but it appears that this integral cannot be evaluated exactly. We shall obtain an approximation for large values of n, by obtaining approximations to λ, $\bar{f}$, and their derivatives, when n is large.

The range of periods, $\dfrac{2\pi}{k}$, will be considered to vary from quantities greater than one up to half the range, that is $\frac{1}{2}(n-1)$. This is chosen on the grounds that the intervals of time would be adjusted so that there would be no expectation of periods less than the interval, and that enough observations would be chosen to include at least two periods in the range. Although this supposes some a priori knowledge of the possible periods, it seems reasonable to expect that the experimenter would have at least a rough idea of the range of periods which might fit his data before attempting to fit a harmonic curve. This range gives a range of values of k from $4\pi/(n-1)$ to $2\pi(1-v)$ where v is arbitrarily small, but fixed. In all cases the epoch, ϵ, varies from 0 to 2π.

It is readily seen that the surface is traced out only once for this range of values of k, ϵ, so that the problem in its approximate form is reduced to that of the evaluation of the definite integral

$$\int_0^{2\pi}\int_{\frac{4\pi}{n-1}}^{2\pi(1-v)} \sqrt{EG - F^2}\, dk\, d\epsilon.$$

We shall obtain the approximations mentioned above, in the first place excluding from consideration values of k in the neighbourhood of 0, π, 2π, the integrals over small ranges including these values being obtained separately.

If k is not in the neighbourhood of 0, π, 2π, we note that

$$\frac{\sin(\frac{1}{2}nk)\cos[\epsilon + \frac{1}{2}k(n-1)]}{\sin(\frac{1}{2}k)}.$$

is a bounded function of k, the upper bound being independent of k, and at most equal to $|\operatorname{cosec}\frac{1}{2}k_1|$, where k_1 is the angle in the range considered nearest to 0, π, 2π. Similarly the upper bound of

$$\frac{\sin(nk)\cos[2\epsilon + k(n-1)]}{\sin k}$$

is at most $|\operatorname{cosec} k_1|$. Hence as n is increased, we may expand $\lambda/n^{\frac{1}{2}}$ in ascending powers of n^{-1}. For large n, therefore, $\lambda = O(n^{-\frac{1}{2}})$, and is approximately $(2/n)^{\frac{1}{2}}$. Since differentiation with respect to k introduces a multiplying factor n in some of the terms, it follows that this is compensated for by the factor λ^{-2} which occurs in the denominator of the derivative, and we may conclude that $\lambda_k = O(n^{-\frac{1}{2}})$. No such compensating factor n occurs in the numerator of λ_ϵ, and it is therefore of order $(n^{-\frac{3}{2}})$. It may readily be seen without actually evaluating the derivatives, which are very long and unwieldy expressions, that

$$\lambda_{kk} = O(n^{\frac{1}{2}}), \lambda_{\epsilon k} = O(n^{-\frac{1}{2}}), \bar{f}_\epsilon = O(n^{-1}), \bar{f}_k = O(1),$$
$$\bar{f}_{kk} = O(n), \bar{f}_{k\epsilon} = O(1), \bar{f} = O(n^{-1}).$$

We thus see that the term of highest order in $E = \dfrac{(n - 1)n(2n - 1)}{12}\lambda^2$.

The term of highest order in $G = n\lambda^2 - 1$.

The term of highest order in $F = \dfrac{n(n - 1)}{4}\lambda^2$.

These are approximately constant for large n, and are equal to $n^2/3$, 1, $n/2$ to a first order of approximation. Hence

$$\sqrt{EG - F^2} \sim \frac{n}{\sqrt{12}}.$$

The range for k may be broken up as follows:

(a) from $\dfrac{4\pi}{n-1}$ to $\dfrac{\alpha}{n^{\frac{1}{2}}}$, where α is a finite angle, independent of n.

(b) from $\dfrac{\alpha}{n^{\frac{1}{2}}}$ to $\pi - \dfrac{\alpha}{n^{\frac{1}{2}}}$

(c) from $\pi - \dfrac{\alpha}{n^{\frac{1}{2}}}$ to $\pi + \dfrac{\alpha}{n^{\frac{1}{2}}}$

(d) from $\pi + \dfrac{\alpha}{n^{\frac{1}{2}}}$ to $2\pi - \dfrac{\alpha}{n^{\frac{1}{2}}}$

(e) from $2\pi - \dfrac{\alpha}{n^{\frac{1}{2}}}$ to $2\pi(1 - v)$.

The method of procedure will be to show that in ranges (a), (c), (e) the integrand is of order n, and that since the ranges in all three cases are of order $n^{-\frac{1}{2}}$, the values of the integrals in these ranges are $O(n^{\frac{1}{2}})$ which is negligible in comparison with the contributions from (b) and (d), which are $O(n)$.

In (a), $\dfrac{4\pi}{n-1} \leq k \leq \dfrac{\alpha}{n^{\frac{1}{2}}}$, we put $k = \dfrac{\alpha}{n^{1-\delta}}$, $\alpha = 4\pi$, and let δ range from p to $\frac{1}{2}$, where p is a positive quantity defined by the relation $(n - 1) = n^{1-p}$. Then λ, $\bar{f}$, are of orders $n^{-\frac{1}{2}}$ and $n^{-\delta}$ respectively. For this range of values of δ, the orders of the derivatives are:

λ_k	λ_{kk}	λ_ϵ	$\lambda_{k\epsilon}$	$\bar{f}_k$	$\bar{f}_\epsilon$	$\bar{f}$
$n^{\frac{1}{2}-\delta}$	$n^{\frac{3}{2}-\delta}$	$n^{-\frac{1}{2}-\delta}$	$n^{\frac{1}{2}-\delta}$	$n^{1-\delta}$	$n^{-\delta}$	$n^{1-\delta}$

These being decreased, it follows that the order of $\sqrt{EG - F^2}$ is not increased for any positive δ, and $\sqrt{EG - F^2} = O(n)$ as before.

In (c), $\pi - \frac{\alpha}{n^{\frac{1}{2}}} \leq k \leq \pi + \frac{\alpha}{n^{\frac{1}{2}}}$, we put $k = \pi \pm \frac{\alpha}{n^{1-\delta}}$, according as $k \gtrless \pi$, and consider $0 \leq \delta \leq \frac{1}{2}$. The orders of the derivatives are as stated in (a) above for this range. The remainder of the range $\pi - \frac{\alpha}{n} < k < \pi + \frac{\alpha}{n}$ is such that the values of the derivatives are of orders as stated with $\delta = 0$, while $\lambda = O(n^{-\frac{1}{2}})$. Thus $\sqrt{EG - F^2} = O(n)$ throughout.

In (e), $2\pi - \frac{\alpha}{n^{\frac{1}{2}}} \leq k \leq 2\pi(1 - v)$, we put $k = 2\pi - \frac{\alpha}{n^{1-\delta}}$, and consider $0 \leq \delta \leq \frac{1}{2}$. In this range the orders of the derivatives are as in (a). In the remainder of the range, $2\pi - \frac{\alpha}{n} < k \leq 2\pi(1 - v)$, the orders of the derivatives are as in (a) with $\delta = 0$, so that $\sqrt{EG - F^2} = O(n)$.

As the ranges (b) and (d) are not independent of n, it remains to be shown that this fact does not affect the final result. We consider, therefore, $k = \frac{\alpha}{n^{1-\delta}}$ and $k = \pi - \frac{\alpha}{n^{1-\delta}}$ where $\frac{1}{2} < \delta < 1$, and since, as in (O) the second and third terms in the denominator of λ are $O(n^{1-\delta})$ and $O(n^{1-2\delta})$ or $O(n^{-1})$ respectively, $\lambda \sim 1 \Big/ \sqrt{\frac{n}{2}}$, while the derivatives have values as in case (a). Thus, in these ranges, $\sqrt{EG - F^2} \sim \frac{n}{\sqrt{12}}$ throughout. Thus we may conclude in all cases that $\sqrt{EG - F^2} = O(n)$.

$$\text{The surface area} = \int_0^{2\pi} \left[\int_{\frac{4\pi}{n-1}}^{\frac{\alpha}{\sqrt{n}}} + \int_{\frac{\alpha}{\sqrt{n}}}^{\pi - \frac{\alpha}{\sqrt{n}}} + \int_{\pi - \frac{\alpha}{\sqrt{n}}}^{\pi + \frac{\alpha}{\sqrt{n}}} + \int_{\pi + \frac{\alpha}{\sqrt{n}}}^{2\pi - \frac{\alpha}{\sqrt{n}}} + \int_{2\pi - \frac{\alpha}{\sqrt{n}}}^{2\pi(1-v)} \sqrt{EG - F^2}\, dk \right] d\epsilon$$

$$= \int_0^{2\pi} \left[\int_{\frac{\alpha}{\sqrt{n}}}^{\pi - \frac{\alpha}{\sqrt{n}}} + \int_{\pi + \frac{\alpha}{\sqrt{n}}}^{2\pi - \frac{\alpha}{\sqrt{n}}} \sqrt{EG - F^2}\, dk \right] d\epsilon$$

$$+ \int_0^{2\pi} \left[\int_{\frac{4\pi}{n-1}}^{\frac{\alpha}{\sqrt{n}}} + \int_{\pi - \frac{\alpha}{\sqrt{n}}}^{\pi + \frac{\alpha}{\sqrt{n}}} + \int_{2\pi - \frac{\alpha}{\sqrt{n}}}^{2\pi(1-v)} \sqrt{EG - F^2}\, dk \right] d\epsilon.$$

In the first two ranges, $\sqrt{EG - F^2} \sim \frac{n}{\sqrt{12}}$

In the last three ranges, $\sqrt{EG - F^2} = O(n)$ and therefore the integral $= O(n^{\frac{1}{2}})$.

Thus the area is equal to

$$(10)\quad \frac{n}{\sqrt{12}}\, 2\pi\left(2\pi - \frac{4\alpha}{n^{\frac{1}{2}}}\right) + \text{terms of lower order} = \frac{4\pi^2}{\sqrt{12}}\cdot n = \frac{2\pi^2}{3}\sqrt{3}\, n.$$

In the case of fitting a linear regression with 3 independent variates, the distribution of R is well known to be

$$\frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma(\frac{3}{2})\Gamma[\frac{1}{2}(n-4)]}(R^2)^{\frac{1}{2}}(1-R^2)^{\frac{1}{2}(n-6)}\, d(R^2).$$

It may readily be seen by a repetition of the argument used in the first paragraph that this expression could be derived by considering the volume of a tube in spherical space of $(n-2)$ dimensions, in which the base surface is a 2-dimensional unit sphere of area 4π. We are assuming that the first approximation to the volume of a tube is equal to the area of the surface multiplied by a fixed function of θ. If, therefore, we divide this expression by 4π, and take R sufficiently close to 1, or $\theta = \cos^{-1}R$ sufficiently close to zero, we shall obtain the expression by which to multiply the surface area, (10), in order to obtain the first approximation to the frequency function of R.

Using Stirling's approximation, we have

$$\Gamma[\tfrac{1}{2}(n-1)] \sim \sqrt{2\pi}\, e^{-\frac{1}{2}(n-1)}[\tfrac{1}{2}(n-1)]^{\frac{1}{2}(n-1)-\frac{1}{2}}$$

and

$$\Gamma[\tfrac{1}{2}(n-4)] \sim \sqrt{2\pi}\, e^{-\frac{1}{2}(n-4)}[\tfrac{1}{2}(n-4)]^{\frac{1}{2}(n-4)-\frac{1}{2}}.$$

The ratio of these $= e^{-\frac{3}{2}}2^{-\frac{3}{2}}\left(1 + \frac{3}{n-4}\right)^{\frac{1}{2}(n-4)}(n-1)(n-4)^{\frac{1}{2}} \sim 2^{-\frac{3}{2}}n^{\frac{3}{2}}.$

Hence the multiplying constant is approximately $n^{\frac{3}{2}}/\sqrt{2\pi}$. Substituting $R = \cos\theta$ in the frequency function divided by this constant, we obtain $2\cos^2\theta\sin^{n-6}\theta\sin\theta\, d\theta$ giving $2\theta^{n-5}\, d\theta$ as the first approximation.

Hence the approximate frequency function for the quantity θ in the case of periodogram analysis is

$$\frac{n^{\frac{3}{2}}}{\sqrt{2\pi}}\, 2\theta^{n-5}\, d\theta\, \frac{\frac{2\pi^2\sqrt{3n}}{3}}{4\pi} = 2^{-\frac{1}{2}}n^{\frac{3}{2}}\pi^{\frac{1}{2}}3^{-\frac{1}{2}}\theta^{n-5}\, d\theta.$$

Thus the first approximation to the probability that θ should be as great or greater is

$$2^{-\frac{1}{2}}n^{\frac{3}{2}}\pi^{\frac{1}{2}}3^{-\frac{1}{2}}\frac{\theta^{n-4}}{n-4}$$

or

$$n^{\frac{3}{2}}\left(\frac{\pi}{6}\right)^{\frac{1}{2}}\theta^{n-4}$$

approximately.

The approximations which have been introduced have been forced upon us by the limitations of the mathematical machinery involved. It must be admitted that these approximations are not those which the experimenter would choose, for the following obvious reason. If we are testing the null hypothesis that the population correlation is zero, for large values of n the sample correlation will approach its expectation value, namely zero, and we shall in general be interested in values of R which are small, and corresponding values of θ in the neighborhood of $\pi/2$. This situation is not provided for in this investigation. It may be, however, that there exists a large correlation in the population, and that owing to the large number in the sample the value of R calculated is near this value. Provided that this population correlation is sufficiently close to unity, the value of θ will be small enough to apply the distribution obtained above, and in such a case will enable us to reject the null hypothesis when the probability calculated from the distribution is sufficiently small.

UNIVERSITY OF LONDON.