THE PROBABILITY LAW FOR THE INTENSITY OF A TRIAL PERIOD, WITH DATA SUBJECT TO THE GAUSSIAN LAW*

BY E. L. DODD

1. *Introduction.* Making use of Lord Rayleigh's[†] probability law for the resultant of n vibrations of fixed and equal amplitude but of chance phase, Schuster[‡] found a probability law for the ratio of the square-root of the intensity of a trial period to the constant term of the Fourier development, and suggested the use of this as a criterion for the possible fortuitous nature of results apparently supporting the existence of periods under investigation. However, as periodicities in a sequence may be invariant to a change of origin, it seems desirable to have a criterion based upon intensity alone, which is thus invariant. Such a criterion will be derived in this paper. Some criterion—as Schuster pointed out—is indispensable in periodogram analysis. It signifies little that one trial period is more probable than another if all the apparent periodicities could be easily the result of chance. The question, then, is this : What fluctuations in the intensity may be expected when the data are subject to chance,—or, more specifically, to the Gaussian law?

2. *The Intensity of a Period and its Probability Law.* Suppose that the probability $p(x)$ that X_r will take on a value $\leqq x$ is given by

$$(1) \quad p(x) = \frac{h}{\pi^{1/2}} \int_{-\infty}^{x} e^{-h^2(t-b)^2} dt, \quad (r = 1, 2, \cdots, km \; ; \; km = n),$$

---

with "precision" h. In testing for a period extending over k variates, suppose the data written in m rows of k elements each:

$$
\begin{matrix}
X_1 & X_2 & \cdots & X_k \\
\end{matrix}
$$

(2)
$$
\begin{matrix}
X_{k+1} & X_{k+2} & \cdots & X_{2k} \\
\cdot & \cdot & \cdots & \cdot \\
\end{matrix}
$$

(3) Averages $\quad Y_1 \quad Y_2 \quad \cdots \quad Y_k,$

where

$$
Y_j = \frac{1}{m} \sum_{r=0}^{m-1} X_{j+kr}.
$$

Taking $\theta = 2\pi/k$, set

$$
S = S(k) = \sum_{r=0}^{k-1} Y_{r+1} \sin r\theta, \quad C = C(k) = \sum_{r=0}^{k-1} Y_{r+1} \cos r\theta,
$$

and

(4)
$$
J = S^2 + C^2, \quad I = \frac{4}{k^2} J.
$$

This quantity I will be called the intensity because, if, instead of supposing X_r governed by (1), we let $X_r = c \sin(r\theta + \alpha)$, thus making it strictly periodic, then $I = c^2$, when $k > 2$.

We note that J is a quadratic form, with determinant of elements

(5)
$$
a_{rs} = \cos (r - s)\theta, \quad (r, s = 1, 2, \cdots, k).
$$

By an orthogonal transformation, J may be reduced to a sum of squares, whose coefficients are the roots λ of the equation $F(\lambda) = 0$, obtained by subtracting λ from the elements of the main diagonal of the preceding determinant and equating this new determinant to zero. When $k > 2$, there are $k - 2$ zero roots ; since, for any r and θ,

(6) $\quad \cos (r - 2)\theta - 2 \cos \theta \cdot \cos (r - 1)\theta + \cos r\theta = 0.$

Indeed,* to the kth column add the $(k-2)$th column, and subtract the $(k-1)$th column multiplied by $2 \cos \theta$. The kth

*For this suggestion, the author is indebted to Miss Edna McCormick.

column becomes divisible by λ. Then take $r = (k-1)$, $(k-2), \cdots, 3$. But

$$(7) \qquad \left\{ \begin{array}{l} (k-1)\cos^2\theta + (k-2)\cos^2 2\theta + \cdots \\ \\ \qquad\qquad + \cos^2 (k-1)\theta = \dfrac{k(k-2)}{4}, \quad \theta = \dfrac{2\pi}{k}, \end{array} \right.$$

is valid for $k > 2$, and with the aid of this, it can be proved that

$$(8) \qquad F(\lambda) = (-1)^k \left\{ \lambda^k - k\lambda^{k-1} + \frac{k^2}{4}\lambda^{k-2} \right\} = 0,$$

which has $k/2$ as a double root.

In (2), let us suppose the origin changed, and that X_r is now the new value obtained by subtracting b from the original X_r. This leaves S, C, J, and I unchanged, and its effect upon (1) is expressed by setting $b = 0$. Using the corresponding new values of Y_r in (3), then, the probability that $y_r < Y_r < y_r + dy_r$ is

$$(9) \qquad \frac{h_1}{\pi^{1/2}} e^{-h_1^2 y_r^2} dy_r, \qquad\qquad h_1 = hm^{1/2},$$

since by (3), each Y_r is the average of m values of X. By an orthogonal transformation, which does not change ΣY_r^2, we may now by (8) reduce J to

$$(10) \qquad J = \frac{k}{2}U^2 + \frac{k}{2}V^2 = u^2 + v^2,$$

where U and V follow the normal law with the same precision h_1, and u and v with precision $h_1(2/k)^{1/2}$. Then the probability that $z < J < z + dz$ is given by*

$$(11) \qquad h_2^2 e^{-zh_2^2} dz, \qquad\qquad h_2^2 = 2h_1^2/k = 2mh^2/k.$$

To obtain the law for I in (5), we need only replace h_2^2 by

$$(12) \qquad H^2 = h_2^2 k^2/4 = kmh^2/2 = nh^2/2.$$

---

* See Czuber, *Theorie der Beobachtungsfehler*, p. 148. Or Helmert, *Ueber die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler.* Zeitschrift für Mathematik und Physik, vol. 21 (1876), pp. 192–218.

THEOREM. *If each of* $n = km$ *independent variates* X_r *is subject to the Gaussian law* (1) *with precision* h, *and if from the averages* Y_r *of the columns of* X's *in* (2), I *is formed in accordance with* (4) ; *then, when* $k > 2$, *the probability* $P(z)$ *that* $I \geqq z$, *is given by*

$$(13) \qquad\qquad P(z) = e^{-nh^2z/2}.$$

In the case of $k = 2$,

$$(14) \qquad P(z) = \frac{2}{\pi^{1/2}} \int_{z'}^{\infty} e^{-t^2} dt, \qquad\qquad z' = \frac{h(nz)^{1/2}}{2}.$$

For the case of $k = 2$, the result is easily obtained by rotating axes through $45°$, noting that $I = (Y_1 - Y_2)^2$.

3. *Application and Discussion.* The averaging process (3) preserves intact a period extending over k items, but tends to smooth out accidental variations and other periods. But this partial smoothing of accidental variations is by no means obliteration. When the value of I in (4) is computed, we ask : is it highly improbable that so large a value would be found if the data contained only chance fluctuations? For example, suppose that the standard deviation σ for 1000 variates under examination is 5, and by the usual rule we take $h^2 = 1/(2\sigma^2)$ $= 0.02$, $nh^2/2 = 10$. For some particular k, say $k = 8$, suppose that $I = 1$. By (13) the probability that a chance normal distribution would give as large a value as 1 is $P(1) = e^{-10} =$ $0.000,05$. Normality granted, the evidence would support the belief that a period covering approximately 8 items exists in the data. The exact determination of this period is beyond the scope of this paper.

In the proof, it has not been assumed that S and C are independent. Such an assumption would, indeed, have led more quickly to the result (13). A slight plausibility for this assumption arises from the fact that the "expected" or "mean" value of SC is zero, if independently the variates are subject to the same arbitrary probability law with finite second moment.

THE UNIVERSITY OF TEXAS