173

ON THE SIMILARITY OF THE DISTRIBUTIONS FOUND FOR THE TEST OF SIGNIFICANCE IN HARMONIC ANALYSIS, AND IN STEVENS'S PROBLEM IN GEOMETRICAL PROBABILITY

By R. A. FISHER

1. Statement of the two problems

In a recent note W. L. Stevens (1939) has published an ingenious solution of the following problem in geometrical probability:

"On the circumference of a circle of unit length, n arcs, each of length x, are marked off at random. What is the probability that every point of the circle is included in at least one of these arcs?"

The expression for this probability at which he arrives is

$$1-n(1-x)^{n-1}+\frac{n(n-1)}{2}(1-2x)^{n-1}-\ldots\pm\frac{n!}{k!\,(n-k)!}(1-kx)^{n-1},$$

in which k is the greatest integer less than $1/x$.

It will be observed that the solution is strikingly similar to one at which I arrived, in 1929, for the apparently very different problem of testing the significance of the largest of the harmonic components into which a series of observations may be analysed.

If a series $u_1, \ldots, u_{2n+1}$ constitute a random sample from a normally distributed population, the linear functions defined by

$$A = S(a_r u_r), \quad B = S(b_r u_r),$$

in which

$$a_r = \sqrt{\left(\frac{2}{2n+1}\right)}\cos\frac{2\pi pr}{2n+1}, \quad b_r = \sqrt{\left(\frac{2}{2n+1}\right)}\sin\frac{2\pi pr}{2n+1},$$

represent the coefficients of a_r and b_r in the harmonic expansion of u. The contribution of any particular period $2(n+1)/p$ is measured by

$$x = A^2+B^2.$$

Values of p from 1 to n supply n values of x, such that

$$\sum_{p=1}^{n}(x) = S(u-\bar{u})^2.$$

The period showing the highest value of x may be tested for significance by considering the fraction of the total sum of squares for which it accounts: thus, if g is the largest of the fractions

$$x/S(x),$$

it was shown that the probability of obtaining so great, or a greater, value, by chance is

$$P = n(1-g)^{n-1} - \frac{n(n-1)}{2}(1-2g)^{n-1} + \ldots + (-)^{s-1}\frac{n!}{k!\,(n-k)!}(1-kg)^{n-1},$$

where k is the greatest integer less than $1/g$.

In other words, the probability that the largest observed fraction is *less* than x is identical with the probability found in Stevens's problem. It is of some interest to elucidate this curious equivalence of the two problems.

2. The equivalence of the problems

The solution of the problem in harmonic analysis was derived from the simultaneous distribution

$$df = e^{-(x_1+x_2+\ldots+x_n)/c}\,dx_1 \ldots dx_n\,c^{-n};$$

the frequency density is constant over any plane finite region for which $S(x)$ is constant, and $x_r \geqslant 0$, for all values of r. These inequalities bound a series of generalized tetrahedra, so that the problem is equivalent to:

If $g_1, \ldots, g_n$ are n fractions such that $S(g) = 1$, and the frequency element is proportional to

$$dg_1, \ldots, dg_{n-1},$$

what is the probability that the largest fraction is less than any assigned value x? The equivalence of this problem to that of Stevens is now readily demonstrated.

Let g_r stand for the fraction of the circumference of the circle (or other closed contour) through which any arc must be shifted in order to coincide with the next, taken in order round the circumference. Then

$$S(g) = 1.$$

Next, so long as neither g_r nor g_{r-1} is made to be negative, any one may be displaced, in such a way that $g_r + g_{r-1}$ is constant, and the frequency with which its value falls in any element dg_r, within these limits, is proportional to dg_r. Hence for simultaneous variation the frequency element is proportional to

$$dg_1 \ldots dg_{n-1}$$

within the limits $g_r \geqslant 0$, $S(g) = 1$.

Now the probability that the greatest fraction is less than or equal to x, is the probability that each fraction without exception is less than or equal to x, and if each arc is of length x, this is the condition that no gaps occur between two successive arcs.

3. Stevens's extension of the problem

Stevens extends his solution to finding the probability that there shall be i gaps uncovered by the chosen arcs. He obtains

$$f(i) = \frac{n!}{i!\,(n-i)!}\left\{(1-ix)^{n-1} - (n-1)(1-(i+1)x)^{n-1} + \ldots \pm \frac{(n-1)!}{(k-i)!\,(n-k)!}(1-kx)^{n-1}\right\}.$$

The equivalence established in § 2 shows that this expression will give also the probability that just i of the fractions g shall exceed x.

The probability that i or more values shall exceed x is therefore

$$\sum_{j=1}^{k}(1-jx)^{n-1}\sum_{j'=1}^{j}\frac{n!}{j'!\,(n-j')!}\frac{(-)^{j-j'}(n-j')!}{(j-j')!\,(n-j)!}.$$

But
$$\sum_{j'=i}^{j}\frac{(-)^{j-j'}n!}{(n-j)!\,j'!\,(j-j')!}=\frac{n!}{j!\,(n-j)!}\sum_{j'=j}^{i}\frac{(-)^{j-j'}j!}{j'!\,(j-j')!}$$
$$=\frac{n!}{j!\,(n-j)!}\frac{(-)^{j-i}(j-1)!}{(j-i)!\,(i-1)!}$$
$$=(-)^{j-i}\frac{n!}{(n-j)!\,(j-i)!\,(i-1)!}\frac{1}{j}.$$

Hence the probability of i or more gaps is

$$\sum_{j=i}^{k}(-)^{j-i}\frac{n!}{(n-j)!\,(j-i)!\,(i-1)!}\frac{(1-jx)^{n-1}}{j}.$$

When $i=1$, we thus have

$$n\left\{(1-x)^{n-1}-(n-1)\frac{(1-2x)^{n-1}}{2}+\frac{(n-1)(n-2)}{2!}\frac{(1-3x)^{n-1}}{3}-\dots\pm\frac{(n-1)!}{(k-1)!\,(n-k)!}\frac{(1-kx)^{n-1}}{k}\right\};$$

when $i=2$,

$$\frac{n(n-1)}{1!}\left\{\frac{(1-2x)^{n-1}}{2}-(n-2)\frac{(1-3x)^{n-1}}{3}+\dots\pm\frac{(n-2)!}{(k-2)!\,(n-k)!}\frac{(1-kx)^{n-1}}{k}\right\}$$

and so on.

The first of these merely verifies the test of significance for the largest of n fractions. The second may be used in a test whether the second largest is significant, such as might be useful if, when the largest is doubtfully significant, it may still be suspected that the two largest are due to some systematic causes. The second largest fraction would then be equated to x. The other cases in which the second largest component is of interest in the interpretation of a series are more complex, and can only properly be discussed in relation to the particular facts exhibited by the data, and the more plausible hypotheses involving periodic disturbances which happen to be in view. For example, if the two most important periods are adjacent, the evidence for reality will be stronger than would appear if they were tested by the formula above, and the same is true, in some material, when one large contribution pertains to a period nearly half that of another. Special distribution problems arise in considering the tests of significance appropriate in each case.

I am indebted to Mr Stevens for the table of numerical values for testing the significance of the second largest harmonic component, with some extension of the values for the largest harmonic component given in my earlier paper.

SUMMARY

The identity of the solutions obtained for two problems apparently quite unconnected is shown to be no coincidence, but due to their intrinsic equivalence.

The further formulae developed by Stevens are also relevant to one of the cases which may arise when harmonic components other than the greatest suggest the reality of periodic disturbances.

A TABLE OF THE TEST OF SIGNIFICANCE IN HARMONIC ANALYSIS.

W. L. STEVENS.

Five per cent values of g_1 and g_2, the largest and second largest fractions

n	g_1	δ_1	g_2	δ_2
3	0·87090	—	0·43545	—
4	0·76792	—	0·39863	—
5	0·68377	—	0·36704	—
6	0·61615	—	0·34021	—
7	0·56115	—	0·31729	—
8	0·51569	—	0·29751	1
9	0·47749	—	0·28028	3
10	0·44495	—	0·26511	8
15	0·33462	1	0·21016	33
20	0·27040	6	0·17547	52
25	0·22805	8	0·15139	64
30	0·19784	10	0·13360	70
35	0·17513	12	0·11986	74
40	0·15738	14	0·10890	75
45	0·14310	14	0·09993	75
50	0·13135	14	0·09244	75

Approximations to g_1 and g_2 may be found by using only the first terms of the respective series. These approximations are *in excess* of the correct values by amounts δ_1 and δ_2, listed in the table; the approximate values of g_1 and g_2 are therefore $g_1 + \delta_1$ and $g_2 + \delta_2$. The errors are even less for smaller probabilities. At the 1 % level for $n = 50$, they are respectively 1 and 21 in the fifth place. The first term approximation is therefore usually adequate in the test of significance.

REFERENCES

W. L. STEVENS (1939). "Solution to a geometrical problem in probability." *Ann. Eugen., Lond.*, **9**, 315–20.

R. A. FISHER (1929). "Tests of significance in harmonic analysis." *Proc. Roy. Soc.* A, **125**, 54–9.