THE SMALL SAMPLE DISTRIBUTION OF $n\omega_n^2$

By A. W. Marshall

The RAND Corporation

The asymptotic distribution of the statistic

$$n\omega_n^2 = n \int_{-\infty}^{\infty} [s_n(x) - F(x)]^2 \, dF(x), \tag{1}$$

where $S_n(x)$ is the sample cumulative distribution function (CDF), and $F(x)$ the true CDF, is known and tabled [1]. Below are tabled some values of the CDF's of $n\omega_n^2$ for $n = 1$, 2, and 3. Convergence to the asymptotic distribution appears to be extremely rapid.

1. General considerations. It is well-known that: (A) the distribution of $n\omega_n^2$ is distribution free so that it is sufficient to treat the case where $F(x)$ is uniform on the interval [0, 1]; and (B) an equivalent form, especially suitable for computation from the ordered sample $x_1 \leqq x_2 \leqq \cdots \leqq x_n$, is

$$n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - F(x_i)\right]^2 \tag{2}$$

or for the case where $F(x)$ is uniform [0, 1]

$$n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - x_i\right]^2. \tag{3}$$

As was suggested to me several years ago by Oliver Gross (3) clearly shows that the CDF of the $n\omega_n^2$ statistic can be evaluated rather easily for small n. The case $n = 1$ is trivial. For $n = 2$ one must evaluate the area in the intersection of a circle with its center at $x_1 = \frac{1}{4}$, $x_2 = \frac{3}{4}$ and a triangle with vertices at (0, 0), (0, 1), and (1, 1). For $n = 3$ one must evaluate the volume in the intersection of a sphere with center at the point $(\frac{1}{6}, \frac{1}{2}, \frac{5}{6})$ and the tetrahedron with vertices at (0, 0, 0), (0, 0, 1), (0, 1, 1), (1, 1, 1). From (3) one also derives the result that $n\omega_n^2$ has a minimum value of $1/12n$ and a maximum value of $n/3$.

2. Case A: $n = 1$. Since

$$\omega_1^2 = \tfrac{1}{12} + [\tfrac{1}{2} - x_1]^2$$

the CDF of ω_1^2 is

$$F_1(z) = \Pr[\omega_1^2 \leqq z] = \begin{cases} 0, & z < \frac{1}{12}, \\ (4z - \frac{1}{3})^{\frac{1}{2}}, & \frac{1}{12} \leqq z \leqq \frac{1}{3}, \\ 1, & z > \frac{1}{3}. \end{cases}$$

3. Case B: $n = 2$.

$$2\omega_2^2 = \tfrac{1}{24} + [\tfrac{1}{4} - x_1]^2 + [\tfrac{3}{4} - x_2]^2.$$

Received June 25, 1957.

By evaluating the area common to a circle of radius $(z - 1/24)^{\frac{1}{2}}$ with center at $(\frac{1}{4}, \frac{3}{4})$ and the triangle with vertices at (0, 0), (0, 1), and (1, 1), multiplying by two, the CDF $F_2(z)$ of the associated value of $2\,\omega_2^2$ is obtained. The result is:

$$
\begin{array}{ll}
0, & z < \frac{1}{24}, \\
2\pi\left(z - \frac{1}{24}\right), & \frac{1}{24} \leqq z \leqq \frac{5}{48}, \\
\left(z - \frac{1}{24}\right)\left[2\pi - 4 \operatorname{Cos}^{-1} \frac{1}{4}(z - 1/24)^{-\frac{1}{2}} + \frac{(z - 5/48)^{\frac{1}{2}}}{z - 1/24}\right], & \frac{5}{48} < z \leqq \frac{1}{6}, \\
\left(z - \frac{1}{24}\right)\left[\frac{3\pi}{2} - 2 \operatorname{Cos}^{-1} \frac{\frac{1}{4}(1/8)^{\frac{1}{2}} - (z - 1/6)^{\frac{1}{2}}(z - 5/40)^{\frac{1}{2}}}{z - 1/24}\right] & \\
\qquad + \frac{1}{8} + \left[\frac{1}{2}(z - 1/6)\right]^{\frac{1}{2}} + \frac{1}{2}(z - 5/48)^{\frac{1}{2}}, & \frac{1}{6} < z \leqq \frac{2}{3}, \\
1, & z > \frac{2}{3}.
\end{array}
$$

4. Case C: $n = 3$. This is the first complicated case and reduces to the problem of evaluating the volume of the intersection of a sphere of radius

$$(z - 1/36)^{\frac{1}{2}},$$

with its center at $(\frac{1}{6}, \frac{1}{2}, \frac{5}{6})$, and a tetrahedron with vertices at (0, 0, 0), (0, 0, 1), (0, 1, 1), and (1, 1, 1). Whereas in the case $n = 2$ there are five intervals over which $F_2(z)$ is separately defined, when $n = 3$ there are eight: $(-\infty, 1/36)$,

TABLE I

Values of the CDF's *of* $n\omega_n^2$ *for* $n = 1, 2, 3$ *and the asymptotic distribution at selected points*

z	$F_1(z)$	$F_2(z)$	$F_3(z)$	$F(z)$
.11888	.37708	.46692	.47343	.50000
.14663	.50318	.57614	.57683	.60000
.16385	.56751	.63384	.63009	.65000
.18433	.63560	.68842	.68521	.70000
.20939	.71009	.73974	.74191	.75000
.24124	.79475	.79126	.79924	.80000
.28406	.89605	.84515	.85481	.85000
.34730	1.00000	.90296	.90617	.90000
.40520	1.00000	.94007	.93661	.93000
.46136	1.00000	.96554	.95723	.95000
.54885	1.00000	.98968	.97793	.97000
.74346	1.00000	1.00000	.99680	.99000
1.16786	1.00000	1.00000	1.00000	.99900

(1/36, 1/18), (1/18, 1/12), (1/12, 1/9), (1/9, 5/24), (5/24, 11/36), (11/36, 1) and (1, ∞). Partial results for these intervals are as follows, where $\zeta = z - 1/36$: $F_3^{(1)}(z) = 0$; $F_3^{(2)}(z) = 8\pi\zeta^{3/2}$; $F_3^{(3)}(z) = \frac{2}{3}\pi(3z - 1/9)$; $F_3^{(4)}(z) = \frac{2}{3}\pi(3z - 1/9) - 2\pi[4\zeta^{3/2} - 2^{\frac{1}{2}}\zeta + 2^{-\frac{1}{2}}/27] + 6\zeta^{3/2}\, V(1, \zeta^{-\frac{1}{2}}/6)$; $\cdots$; $F_3^{(8)}(z) = 1$, where $V(1, a)$ is the volume of the wedge-shaped segment of the sphere of unit radius, center at the origin, cut out by the two planes $x = a$ and $y = a$. It is possible to obtain expressions in closed form for $F_3(z)$ over all of the eight intervals; however their derivation is tedious and the expressions complicated.[1] A numerical evaluation was therefore undertaken by the RAND Numerical Analysis section. The result of these computations are shown in Table 1 along with the calculated values of $F_n(z)$ for $n = 1$ and 2, and for the asymptotic distribution. The values of $F_3(z)$ appear to be off by one in the fifth decimal place. The rapid convergence to the asymptotic distribution, especially in the more interesting region of the tail of distribution, seems clear.

One other piece of evidence, although of a much weaker sort, is available that suggests that the asymptotic distribution is a good approximation to the exact distribution for small n. A sample of 400 values of $n\omega_n^2$ was produced for the case $n = 10$. Grouping into twenty cells using the 5th, 10th, 15th, $\cdots$, percentage points of the asymptotic distribution gave the following cell entries: 13, 19, 20, 18, 11, 21, 16, 18, 28, 17, 21, 22, 26, 18, 21, 16, 23, 25, 23, 24. Application of the χ^2 test gives a value of $\chi^2 = 17.5$. With 19 d.f. this value is exceeded with probability of approximately .55. Application of the Kolmogorov test statistic, $\text{Sup} \mid S_n(x) - F(x) \mid$, to the grouped data (for an approximate test) gives a value of 1.20. This value would be exceeded on the order of 11 per cent of the time under the null hypothesis.

Anderson and Darling in one of their papers [2] mention that "empirical study suggests that the asymptotic value is reached very rapidly, and it appears safe to use the asymptotic value for a sample size as large as 40." The results given above suggest the sample size for which it is reasonable to use the asymptotic distribution is likely to be more nearly 3 or 4, or perhaps 5.

For an allied form of the ω^2 test criterion, denoted by W_n^2 in [2] and formed by adding to (1) the weight function $\psi(X) = [F(X)(1 - F(X))]^{-1}$, an even more rapid convergence seems to occur. $F_1(z) = (1 - 4e^{-z-1})^{1/2}$ for the statistic W_1^2. Evaluating $F_1(z)$ at the 90, 95, and 99 asymptotic percentage points given in [2] yields .88716, .93292, and .98433 respectively.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 765–769.

[1] D. Anderson may publish the complete results later.