SIMPLE APPROXIMATIONS TO THE PROBABILITY INTEGRAL AND $P(\chi^2, 1)$ WHEN BOTH ARE SMALL

By J. B. S. HALDANE
Indian Statistical Institute

The most frequently used test of significance involves either the "tail" of the probability integral, or what is equivalent, the probability of χ^2 with one degree of freedom. I believe, moreover, that it is more efficient, when using tests of significance, to think in terms of $-\log Q$ (in the Pearson-Hartley 1954) notation, rather than Q. If $-\log Q < 1.3$ we do not usually regard a deviation as significant. If $-\log Q > 3$ we regard it as highly significant. Good (1950) has given an approximation to $-\log Q$, for a reduced normal variate, namely

$$-10 \log Q = 2\tfrac{1}{6} x^2+4+10 \log_{10} x. \qquad \ldots \quad (1)$$

which is adequate for most purposes. My own is somewhat more accurate. It is well known that for a reduced normal distribution

$$Q = \int_x^\infty (2\pi e^{t^2})^{-\frac{1}{2}} dt$$
$$= (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} x^{-1}(1-x^{-2}+1.3.x^{-4}-1.3.5.x^{-6}+1.3.5.7.x^{-8}-),$$

the series being of course an asymptotic expansion which is of little value for computation till $x > 4$. Hence

$$Q = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}(x^2+2-3x^{-2}+16x^{-4}-124x^{-6}+1224x^{-8}+\cdots)^{-\frac{1}{2}}$$
$$\ln Q = -\tfrac{1}{2}x^2-\tfrac{1}{2}\ln(x^2+2)-\tfrac{1}{2}\ln 2\pi-\tfrac{1}{2}\ln(1-3x^{-4}+22x^{-6}-168x^{-8}+1560x^{-10}-)$$
$$= -\tfrac{1}{2}x^2-\tfrac{1}{2}\ln(x^2+2)-\tfrac{1}{2}\ln 2\pi+\tfrac{1}{2}(3x^{-4}-22x^{-6}+\frac{345x^{-8}}{2}-1626x^{-10}+)$$

$$-\log_{10} Q = \frac{1.52x^2}{7}+\tfrac{1}{2} \log_{10}(x^2+2)+0.399 \text{ nearly.} \qquad \ldots \quad (2)$$

I use the approximation $\frac{3.04}{7} = .434286$ for $\log_{10} e = .434294$. Good's approximation is equivalent to $.4\dot{3}$. The following table shows the accuracy of (2).

x	2	2.5	3	3.5
$-\log Q$ from (2)	1.657	2.214	2.874	3.636
$-\log Q$ (correct)	1.643	2.207	2.870	3.633

It is clear that for values of x exceeding 2 the approximation is quite satisfactory, and there is no objection to using .40 instead of .399. Curiously enough the expression

$$\frac{1.52x^2}{7}+\tfrac{1}{2} \log(x^2+2-3x^{-2})+.399$$

does not give any greater accuracy in the neighbourhood of 2. When $x = 2$, it gives $-\log Q = 1.628$, whose error happens to be the same as that of (2).

For χ_1^2 we have

$$Q(\chi^2|1) = \frac{1.52\chi^2}{7} + \tfrac{1}{2}\log_{10}(\chi^2+2) + .098 \quad \text{nearly.} \qquad \dots \quad (3)$$

We may use 0.1 for .098 with little disadvantage. Clearly the error is the same as that of (2).

A more interesting, but probably less useful, approximation to the asymptotic expansion is obtained as follows.

$$\begin{aligned} & 1-x^{-2}+3x^{-4}-15x^{-6}+105x^{-8}-945x^{-10}+10395x^{-12}-135135x^{-14} \\ & = 1-(x^2+3)^{-1}-6x^{-6}(1-13x^{-2}+144x^{-4}-1692x^{-6}+22401x^{-8}-) \\ & = 1-(x^2+3)^{-1}-6(x^6+13x^4+25x^2+145)^{-1}-12720x^{-14}+\cdots. \end{aligned}$$

Unfortunately the next polynomial, if the process is repeated, does not have integral coefficients. However the expression

$$\int_x^\infty (2\pi e^{t^2})^{-\frac{1}{2}}dt = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}x^{-1}[1-(x^2+3)^{-1}-6(x^6+13x^4+25x^2+145)^{-1}+O(x^{-14})] \qquad \dots \quad (4)$$

is more accurate than (2) for $x \geqslant 2$. When $x = 2$ it gives $-\log Q = 1.6411$, or $Q = .022826$, the correct value being .022750. For values of x exceeding 3 it is very accurate, and might perhaps have been used instead of the continued fraction which Sheppard (1939) actually employed. Its greatest interest is perhaps that it can be employed for the approximate summation of a number of asymptotic expansions, such as that for the "tail" of the exponential integral. The general expression is

$$\begin{aligned} & 1-ht^{-1}+h(h+1)t^{-2}-h(h+1)(h+2)t^{-3}+\dots \\ & = 1-h(t+h+1)^{-1}-h(h+1)[t^3+(3h+5)t^2+(h+2)(3h+1)t+(h+2)(h^2+7)]^{-1}+O(t^{-7}). \end{aligned} \qquad \dots \quad (5)$$

On putting $h = \frac{1}{2}$, $t = \frac{1}{2}x^2$, we readily find (4).

There are sound theoretical grounds for using $\log\left(\frac{Q}{1-Q}\right)$ as a measure of the credibility of a hypothesis. It is however hard to find a simple approximation to this quantity. I suggest that Q might be expressed in C.W. Allen's (1955) expression *dex*, (n *dex* $= 10^n$). Q^{-1} might be called the improbability. Thus if on a given hypothesis $P(\chi^2)$ were .02 for a particular sample, we might say that on the basis of this sample the improbability of the hypothesis was 1.7 *dex*.

REFERENCES

ALLEN C. W. (1955): *Astrophysical Quantities*, Athlone Press, London.

GOOD, I. J. (1950): *Probability and the Weighing of Evidence*, Griffin, London.

PEARSON, E. S., and HARTLEY, H. O. (1954): *Biometrika Tables for Statisticians*, Vol. 1. Cambridge University Press.

SHEPPARD W. F. (1939): The probability integral. *Brit. Ass. Math. Tab.*, 7, Cambridge University Press.

Paper received : April 1960.