THE COMPUTATION OF THE X-DISTRIBUTION

B. L. VAN DER WAERDEN
UNIVERSITY OF ZURICH

The X-test is a two-sample test, defined as follows. Let $x_1, \cdots, x_g$ and $y_1, \cdots, y_h$ be independent observed variables. Let $r_1, \cdots, r_g$ be the rank numbers of $x_1, \cdots, x_g$ among the x's and y's. Put $g + h = n$. Let Φ be the (cumulative) normal distribution function and $\Psi = \Phi^{-1}$ the inverse function. Put

$$a_r = \Psi\left(\frac{r}{n+1}\right), \qquad r = 1, \cdots, n. \tag{1}$$

The hypothesis H to be tested is: The x's have the same distribution as the y's. The test statistic is

$$X = \sum a_r, \tag{2}$$

the summation extending over the rank numbers $r_1, \cdots, r_g$ of the x's. If X exceeds a limit X_β depending on the level β, the hypothesis H is rejected. The two-sided test on the level 2β rejects when the absolute value $|X|$ exceeds the same limit X_β.

In my paper [1] I have proved that under the hypothesis H the statistic X is asymptotically normal for $g/h \to \infty$ or $h/g \to \infty$. Noether, in his review of my paper [2], pointed out that the asymptotic normality for $g + h \to \infty$ can also be proved when g/h and h/g remain bounded. A full proof for $g \to \infty$ and $h \to \infty$ was given by D. J. Stoker in his Amsterdam thesis [3].

For small g and h the exact limit X_β can be found by explicit computation of the largest X-values. Beyond $g = h = 10$, this computation becomes impracticable. The normal distribution may be used as an approximation, but the comparison with the exact values for $g = h = 8$ or 9 or 10 showed a systematic deviation. The normal approximation for X_β was always too large, so that the power of the test was diminished.

A closer examination showed that this deviation is mainly due to the rather large terms a_1 and a_n, which may or may not be included in the sum (2). An improved approximation could be obtained by separating these large terms from the sum (2).

Consider, for example, the case $g = h = 5$. The 10 terms a_r are, according to (1),

$$\begin{array}{lllll} a_1 = -1.34 & a_2 = -.91 & a_3 = -.60 & a_4 = -.35 & a_5 = -.11 \\ a_6 = +.11 & a_7 = +.35 & a_8 = +.60 & a_9 = +.91 & a_{10} = +1.34\,. \end{array} \tag{3}$$

The test statistic X is a sum of $g = 5$ terms a_r chosen at random from the 10 possible terms (3). Now if X were a sum of many terms, each having only a relatively small influence, the normal approximation would be very good. However, the terms a_1 and a_{10} are not small. Therefore they have to be considered separately.