REFERENCES

[1] P. S. Dwyer, *Linear Computations*, John Wiley and Sons, 1951.
[2] P. S. Dwyer, "The square root method and its use in correlation and regression," *J. Amer. Stat. Assn.*, Vol. 40 (1945), pp. 493–503.
[3] J. P. Gram, "Ueber die entwicklung reeler functionen in reihen," Crelle's *J. für Math.*, Vol. 94 (1883), pp. 41–73. (This is a presentation in German of his 1879 dissertation.)
[4] A. S. Householder, "Some numerical methods for solving systems of linear equations," *Amer. Math. Monthly*, Vol. 57 (1950), pp. 453–459.
[5] B. F. Kimball, "Orthogonal polynomials applied to least square fitting of weighted observations," *Ann. Math. Stat.*, Vol. 11 (1940), p. 352.

ON THE ASYMPTOTIC NORMALITY OF CERTAIN RANK ORDER STATISTICS[1]

By Meyer Dwass

Northwestern University

1. Summary. Let $(R_1, \cdots, R_N)$ be a random vector which takes on each of the $N!$ permutations of the numbers $(1, \cdots, N)$ with equal probability, $1/N!$. Sufficient conditions are given for the asymptotic normality of $S_N = \sum_{i=1}^N a_{Ni} b_{NR_i}$, where $(a_{N1}, \cdots, a_{NN})$, $(b_{N1}, \cdots, b_{NN})$ are two sets of real numbers given for every N. These sufficient conditions are apparently quite different from those given by Wald and Wolfowitz [9] and extended by various writers [4, 7]. In some situations the conditions given here may be easier to apply than those given previously. The most general conditions available to date appear to be those of Hoeffding [4]. In the examples below, however, is given a case of an S_N which does not satisfy the conditions required by Hoeffding's theorem but which is asymptotically normal by our results.

2. Statement of theorem and its proof. We will assume hereafter that

$$\sum_{i=1}^N a_{Ni} = \sum_{i=1}^N b_{Ni} = 0, \qquad \sum_{i=1}^N a_{Ni}^2 = 1.$$

Theorem. Suppose for an integer $k \geqq 1$ there is a random variable X satisfying the following conditions:

(a) X has a continuous cdf $F(x)$,

(b) if $X_1, \cdots, X_N$ are independent random variables each with the cdf $F(x)$ and $Z_{N1} \leqq \cdots \leqq Z_{NN}$ are the ordered values of $X_1, \cdots, X_N$ then

$$b_{Ni} = EZ_{Ni}^k - \sum_{j=1}^N EZ_{Nj}^k / N$$

for all N and i.

Received 7/21/52, revised 11/24/52.

[1] Work done under the sponsorship of the Office of Naval Research while the author was a graduate student at the University of North Carolina.

(c) $E\,|\,X\,|^{3k} < \infty$.

(d) Either X_i^k is normal or (e) $\max_{1 \leqq i \leqq N} |\, a_{Ni} \,| \to 0$ as $N \to \infty$.

Then S_N is asymptotically normally distributed.

PROOF OF THEOREM. Associate with the random vector $X_1, \cdots, X_N$ the random vector $R_1, \cdots, R_N$ where R_i = number of $X_j \leqq X_i$.

Let $g_N(X) = g_N(X_1, \cdots, X_N)$ be the random variable $g_N(X) = \sum_{i=1}^N a_{Ni} b_{NR_i}$. Hence, for every N, the distribution of $g_N(X)$ is identical with that of S_N, for each assumes the same set of values with the same probabilities. Write

$$g_N(X) = \sum_{i=1}^N a_{Ni} X_i^k - \left(\sum_{i=1}^N a_{Ni} X_i^k - g_N(X)\right).$$

If it can be shown that

$$\text{(1)} \qquad \sum_{i=1}^N a_{Ni} X_i^k - g_N(X)$$

converges in probability to zero, then if $\sum_{i=1}^N a_{Ni} X_i^k$ has a limiting distribution, $g_N(X)$ will approach that same limiting distribution (as $N \to \infty$) ([1], p. 254).

That $\sum_{i=1}^N a_{Ni} X_i^k$ has a limiting normal (0, 1) distribution is seen by applying the condition of Liapounoff that

$$\frac{\left(\sum_{i=1}^N |\, a_{Ni} \,|^3 E \,|\, X^k - EX^k \,|^3\right)^{\frac{1}{3}}}{(E(X^k - EX^k)^2)^{\frac{1}{2}}} \to 0$$

as $N \to \infty$. This is so, since

$$\sum_{i=1}^N |\, a_{Ni} \,|^3 \leq \max_{1 \leqq i \leqq N} |\, a_{Ni} \,| \sum_{j=1}^N (a_{Nj})^2 = \max_{1 \leqq i \leqq N} |\, a_{Ni} \,|.$$

To show that (1) converges in probability to zero, it will be sufficient to show that $\lim_{N\to\infty} E(\sum_{i=1}^N a_{Ni} X_i^k - g_N(X))^2 = 0$. Denote by U_N the expression

$$\begin{aligned} U_N &= E\left(\sum_{i=1}^N a_{Ni} X_i^k - g_N(X)\right)^2 = E\left(\sum_{i=1}^N a_{Ni}(X_i^k - EX_i^k) - g_N(X)\right)^2 \\ &= E(X^k - EX^k)^2 - \frac{2}{N!} \sum{}' \left[\left(\int N! \sum_{i=1}^N a_{Ni}(X_i^k - EX_i^k) \prod_{i=1}^N dF(x_i)\right) \right. \\ &\qquad \left. \cdot \sum a_{Ni} b_{Nr_i}\right] + Eg_N^2(X) \end{aligned}$$

where the integral is over that part of the space where $R_i = r_i$ ($i = 1, \cdots, N$) and $r_1, \cdots, r_N$ is one of the $N!$ permutations of $1, \cdots, N$ and where the summation $\sum'$ is over all such permutations.

By condition (b) and by the fact that $N^{-1} \sum_{i=1}^N EZ_{Ni}^k = EX^k$, it follows that $U_N = E(X^k - EX^k)^2 - Eg_N^2(X)$. By straightforward algebra,

$$Eg_N^2(X) = \frac{1}{N!} \sum{}' \left(\sum_{i=1}^N a_{Ni}\, b_{Nr_i}\right)^2 = \frac{1}{N-1} \sum_{i=1}^N b_{Ni}^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}(EZ_{Ni}^k)^2 - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} EZ_{Ni}^k\right)^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}(EZ_{Ni}^k)^2 - \frac{N}{N-1}(EX_k)^2.$$

By a theorem of Hoeffding [3]

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(EZ_{Ni}^k)^2 = EX^{2k} \qquad \text{for } k \geqq 1. \tag{2}$$

Hence $\lim_{N\to\infty} U_N = 0$, which proves the theorem.

3. Applications.

EXAMPLE 1. Consider the test studied by Hotelling and Pabst [5] based on the statistic $S_N' = \sum_{i=1}^N iR_i$. This statistic was shown to be asymptotically normal in [5]. If we set $a_{Ni} = (i - (N+1)/2)/N(N^2-1)/12$ and $b_{Ni} = i/(N+1) - 1/2$, then it is easy to see that the random variable X which has uniform distribution on the unit interval satisfies the conditions of the theorem with $k = 1$. Hence S_N is asymptotically normal and therefore so is S_N'.

EXAMPLE 2. The statistic $S_N = \sum_{i=1}^N a_{Ni}EZ_{NR_i}$, where the Z_{Ni} are order statistics from a normal (0, 1) population and the a_{Ni} satisfy certain conditions, has been studied by Hoeffding and others [8] and shown to be asymptotically normal. Our theorem shows S_N to be asymptotically normal not only for the case of normal order statistics but also when the Z_{Ni} are order statistics from any population satisfying conditions (a), (c) and (e). The last will be satisfied, for instance, when

$$a_{Ni} = \begin{cases} \sqrt{n/(mN)} & (i = 1, \cdots, m) \\ -\sqrt{m/(mN)} & (i = m+1, \cdots, m+n), \end{cases} \tag{3}$$

where $m + n = N$ and m and n both approach infinity as N approaches infinity. This type of a_{Ni} is commonly used in the "two-sample problem."

EXAMPLE 3. When $a_{Ni}[\sum_{i=1}^N (EZ_{Ni} - \sum_{i=1}^N EZ_{Ni}/N)^2]^{\frac{1}{2}} = EZ_{Ni} - \sum_{i=1}^N EZ_{Ni}/N$ and $b_{Ni} = EZ_{Ni} - \sum_{i=1}^N EZ_{Ni}/N$, this S_N has been studied by Hoeffding [2] for the case of Z_{Ni} from a normal (0, 1) population. In this case he showed S_N to be asymptotically normal. Our theorem shows this is also true when the Z_{Ni} are order statistics from any population satisfying (a) and (c), $(k = 1)$, since (e) holds. This is so since $\max_{1\leqq i\leqq N} |a_{Ni}|$ is given for either the index 1 or N. Assume it is N. We have $EZ_{NN}^j = N\int_{-\infty}^{\infty} x^j F^{N-1}(x)dF(x)$, $(j = 1, 2)$, and an easy argument gives that $\lim_{N\to\infty} EZ_{NN}^j/N = 0$. This and the fact that $(EZ_{NN})^2 \leqq EZ_{NN}^2$ together with (2) proves the assertion. If the index is 1, the proof is analogous.

EXAMPLE 4. When the a_{Ni} are given by (3) and $b_{Ni} = i/(N+1) - \frac{1}{2}$ the statistic S_N is, for every N, linearly related to the Wilcoxon statistic, further discussed by Mann and Whitney [6], which, as is well known, is asymptotically normal. This is also seen from our theorem for reasons stated in Examples 1 and 2.

EXAMPLE 5. In a thesis by Terry [8], the statistic $m - \sum_{i=1}^{m} EZ_{NR_i}^2$ (where the Z_{Ni} are the order statistics from a normal (0, 1) population) is proposed against the alternative that the X_i are normal with common mean, the first m having one variance, the remaining $M - n$ another. This statistic is linearly related to an S_N where the a_{Ni} are given by (3) and $b_{Ni} = EZ_{Ni}^2 - \sum_{j=1}^{N} EZ_{Nj}^2/N$. No consideration of the asymptotic distribution of this statistic is made in [8]. We see that this S_N is asymptotically normal when the Z_{Ni} are order statistics from any population satisfying (a) and (c).

By way of example of a case not covered by earlier theorems (for instance, see Theorem 4 of [4]) we take $S_N = \sum_{i=1}^{N} a_{Ni}EZ_{NR_i}$ where the Z_{Ni} are order statistics from a normal (0, 1) population and where condition (13) of [4] is not satisfied. We can construct such a case as follows. Let the a_{Ni} be given by (3) but let the integer m be fixed and independent of N. Then condition (13) of [4] says that

$$\left[n\left(\frac{m}{nN}\right)^{r/2} + m\left(\frac{-n}{mN}\right)^{r/2}\right]\frac{\sum_{i=1}^{N} EZ_{Ni}^{r}/N}{\left[\sum_{i=1}^{N} EZ_{Ni}^2/N\right]^{r/2}} \tag{4}$$

must approach zero as N approaches infinity for $r = 3, 4, \cdots$. From [3] we have that $\sum_{i=1}^{N} EZ_{Ni}^{j}/N$ has for its limit the jth moment of a normal (0, 1) variable. Hence for even r, (4) does not approach zero. However, we see from our theorem that S_N is asymptotically normal.

REFERENCES

[1] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] W. HOEFFDING, "Most powerful rank order tests of independence," unpublished paper, 1950.

[3] W. HOEFFDING, "On the distribution of the expected values of order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 93–100.

[4] W. HOEFFDING, "A combinatorial central limit theorem," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 558–566.

[5] H. HOTELLING AND M. R. PABST, "Rank correlation and tests involving no assumption of normality," *Ann. Math. Stat.*, Vol. 7 (1936), pp. 29–43.

[6] H. B. MANN AND R. R. WHITNEY, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 50–60.

[7] G. E. NOETHER, "On a theorem by Wald and Wolfowitz," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 455–458.

[8] M. TERRY, "Some rank order tests which are most powerful against specific parametric alternatives," Univeristy of North Carolina Library Unpublished dissertation, 1951.

[9] A. WALD AND J. WOLFOWITZ, "Statistical tests based on permutations of the observations," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 358–372.