

These recursion formulas may be obtained directly by equating the coefficients of x^r in

$$(11a) \quad f(x; k, m) = f(x; 1, m)f(x; k - 1, m),$$

$$(11b) \quad f(x; 1, m)f'(x; k, m) = k \cdot f(x; k, m)f'(x; 1, m),$$

or

$$(11c) \quad (1 - x)f(x; k, m) = (1 - x^{m+1})f(x; k - 1, m).$$

Equation (10c) may also be obtained as an immediate consequence of (10a).

Suitable tables for an exact test of significance for $m = 2$ are in preparation. Tables of the restricted-occupancy coefficients $N(r, k, 4)$ for values of k from 1 to 20 and of r from 1 to 40 are given in [4]. It was also found experimentally for $m = 2$ and $m = 4$ that the chi-square criterion applied to the observed v_i and their expectations provides a good approximate test of significance.

REFERENCES

- [1] E. BATICLE, "Le problème de la répartition," *C. R. Acad. Sci. Paris*, Vol. 197 (1933), pp 632-634.
- [2] E. BATICLE, "Le problème de la répartition," *C. R. Acad. Sci. Paris*, Vol. 201 (1935), pp. 862-864.
- [3] E. BATICLE, "Le problème des stocks," *C. R. Acad. Sci. Paris*, Vol. 222 (1946), pp. 355-357.
- [4] A. N. POZNER, "A Restricted Occupancy Criterion for Tests of Consistency and Inconsistency in Subjective Ratings," M. A. thesis, Virginia Polytechnic Institute, 1955.

A VECTOR FORM OF THE WALD-WOLFOWITZ-HOEFFDING THEOREM¹

BY D. A. S. FRASER

University of Toronto and Princeton University

1. Summary. Hotelling and Pabst [1] showed that the rank correlation coefficient had a limiting normal distribution under the equally likely permutations of the hypothesis of independence. Wald and Wolfowitz [2] developed a general theorem of this type, and Noether [3] and Hoeffding [4] have relaxed the conditions used therein. In this paper a vector form of the theorem is proved along the lines used in an example by Wald and Wolfowitz [1] but taking account of the singular cases in which the correlations approach one.

2. The theorem. For each positive integer n let $\|C_{n1}(i, j)\|, \dots, \|C_{nk}(i, j)\|$ be $n \times n$ matrices of real numbers. Also let (R_1, \dots, R_n) be a random variable which takes each permutation of $(1, \dots, n)$ with the same probability, $(n!)^{-1}$.

Received February 18, 1955.

¹ Prepared in connection with research sponsored by the Office of Naval Research at Princeton University.

We shall be concerned with the limiting form of the joint distribution of S_{n1}, \dots, S_{nk} , where

$$S_{n1} = \sum_{i=1}^n C_{n1}(i, R_i),$$

.....

$$S_{nk} = \sum_{i=1}^n C_{nk}(i, R_i).$$

First we find the means, variances, and covariances. Obviously

$$(1) \quad E\{S_{n\alpha}\} = \frac{1}{n} \sum_{i,j=1}^n C_{n\alpha}(i, j).$$

For the variances and covariances, it is convenient to adjust the matrices so that they sum to zero by rows and columns; we define

$$(2) \quad d_{n\alpha}(i, j) = C_{n\alpha}(i, j) - \frac{1}{n} \sum_{k=1}^n C_{n\alpha}(k, j) - \frac{1}{n} \sum_{i=1}^n C_{n\alpha}(i, 1) + \frac{1}{n^2} \sum_{k,l=1}^n C_{n\alpha}(k, l),$$

and notice that

$$(3) \quad S_{n\alpha} - E\{S_{n\alpha}\} = \sum_{i=1}^n d_{n\alpha}(i, R_i).$$

From the equations

$$\sum_{i=1}^n d_{n\alpha}(i, j) = 0 = \sum_{j=1}^n d_{n\alpha}(i, j),$$

it is straightforward to prove that

$$(4) \quad \text{cov} \{S_{n\alpha}, S_{n\beta}\} = \frac{1}{n-1} \sum_{i,j=1}^n d_{n\alpha}(i, j) d_{n\beta}(i, j).$$

And in particular

$$(5) \quad \text{var} \{S_{n\alpha}\} = \frac{1}{n-1} \sum_{i,j=1}^n d_{n\alpha}^2(i, j).$$

We designate by $\rho_{n\alpha\beta}$ the correlation between $S_{n\alpha}$ and $S_{n\beta}$; then we have

THEOREM. *If for each $\alpha = 1, \dots, k$*

$$(6) \quad \lim_{n \rightarrow \infty} \frac{\frac{1}{n} \sum_{i,j=1}^n d_{n\alpha}^r(i, j)}{\left[\frac{1}{n} \sum_{i,j=1}^n d_{n\alpha}^2(i, j) \right]^{r/2}} = 0, \quad r = 3, 4, \dots,$$

then a necessary and sufficient condition that

$$\left(\frac{S_{n1} - E\{S_{n1}\}}{[\text{Var} \{S_{n1}\}]^{1/2}}, \dots, \frac{S_{nk} - E\{S_{nk}\}}{[\text{Var} \{S_{nk}\}]^{1/2}} \right)$$

have a limiting k -variate normal distribution is that the correlations $\rho_{n\alpha\beta}$ approach limits as $n \rightarrow \infty$. The limiting distribution has means zero, variances one, and correlations $\rho_{\alpha\beta}$, where

$$\rho_{\alpha\beta} = \lim_{n \rightarrow \infty} \rho_{n\alpha\beta}.$$

Condition (6) is satisfied if for all $\alpha = 1, \dots, k$

$$(7) \quad \lim_{n \rightarrow \infty} \frac{\max_{i,j=1}^n d_{n\alpha}^2(i, j)}{\frac{1}{n} \sum_{i,j=1}^n d_{n\alpha}^2(i, j)} = 0.$$

PROOF. The proof is based on Hoeffding's Theorem 3, which is essentially the theorem above with $k = 1$. We assume, without loss of generality, that for each α and for each n large enough that the denominator of (7) is nonzero, a scale factor has been applied to the $C_{n\alpha}(i, j)$, $d_{n\alpha}(i, j)$ to make $\text{Var} \{S_{n\alpha}\} = 1$. Also for simplicity we consider the case $k = 2$.

Consider a linear combination, $\hat{d}_n(i, j) = ad_{n1}(i, j) + bd_{n2}(i, j)$, and the corresponding random variable

$$\begin{aligned} \hat{S}_n - E\{\hat{S}_n\} &= \sum_{i,j=1}^n \hat{d}_n(i, R_i) \\ &= a[S_{n1} - E\{S_{n1}\}] + b[S_{n2} - E\{S_{n2}\}]. \end{aligned}$$

We check to see if this random variable satisfies the conditions of Hoeffding's theorem. We have

$$\begin{aligned} \frac{1}{n} \sum_{i,j=1}^n |\hat{d}_n(i, j)|^r &\leq \max \{ |a|^r, |b|^r \} \frac{1}{n} \sum_{i,j=1}^n [|d_{n1}(i, j)| + |d_{n2}(i, j)|]^r \\ &\leq \max \{ |a|^r, |b|^r \} \frac{1}{n} \sum_{i,j=1}^n [|2d_{n1}(i, j)|^r + |2d_{n2}(i, j)|^r] \\ &\leq 2^r \max \{ |a|^r, |b|^r \} \left[\frac{1}{n} \sum_{i,j=1}^n |d_{n1}(i, j)|^r + \frac{1}{n} \sum_{i,j=1}^n |d_{n2}(i, j)|^r \right]. \end{aligned}$$

Now if $\lim \rho_{n12} (= \rho_{12})$ exists, then the limit of $\text{Var} \hat{S}_n$ exists. If this limit is zero, then $\hat{S}_n - E\{\hat{S}_n\}$ converges in probability to zero; that is, it has a degenerate normal limiting distribution. If the limit of $\text{Var} \{ \hat{S}_n \}$ is greater than zero, then (8) implies that Condition (6) is satisfied. However, for this it is necessary to note that Condition (6) is equivalently given by the same expression but with absolute bars around $d_{n\alpha}(i, j)$; this was proved by Hoeffding in [4]. It follows then that

$$\hat{S}_n - E\{\hat{S}_n\} = a[S_{n1} - E\{S_{n1}\}] + b[S_{n2} - E\{S_{n2}\}]$$

has a limiting normal distribution. This limiting distribution has mean zero and variance $a^2 + b^2 + 2ab\rho_{12}$.

Let (Y, Y') designate a random variable having a bivariate normal distribution with means 0, variances 1, and correlation ρ_{12} . Then the limiting distribution of $a(S_{n1} - E\{S_{n1}\}) + b(S_{n2} - E\{S_{n2}\})$ is the distribution of $aY + bY'$ for all a, b . If we knew that $(S_{n1} - E\{S_{n1}\}, S_{n2} - E\{S_{n2}\})$ had a limiting distribution, say the distribution of a random variable (Z, Z') , then it would follow that the linear compound would have the distribution of $aZ + bZ'$. But this means that the random variable $aY + bY'$ is equivalent to $aZ + bZ'$ for all a, b . By Cramér [5], p. 105, this implies that the random variables (Z, Z') and (Y, Y') are equivalent. If a limiting distribution did not exist for

$$(S_{n1} - E\{S_{n1}\}, S_{n2} - E\{S_{n2}\}),$$

then we could extract on n two subsequences which have limiting distributions that are different. This contradicts the statement that the limiting distribution must be that of (Y, Y') . This proves the limiting normality when the correlation approaches a limit.

If $\lim_{n \rightarrow \infty} \rho_{n12}$ does not exist, then we can extract on n two subsequences with different limits. Then, by the argument above, the two subsequences of random variables would have limiting normal distributions which are different. This implies that the original sequence of random variables does not have a limiting distribution. The proof is completed.

REFERENCES

- [1] H. HOTELLING AND M. PABST, "Rank correlation and tests of significance involving no assumption of normality," *Ann. Math. Stat.*, Vol. 7 (1936), pp. 29-43.
- [2] A. WALD AND J. WOLFOWITZ, "Statistical tests based on permutations of the observations," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 358-372.
- [3] G. E. NOETHER, "On a theorem by Wald and Wolfowitz," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 455-458.
- [4] W. HOEFFDING, "A combinatorial central limit theorem," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 558-566.
- [5] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge University Press, 1937.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Princeton meeting of the Institute, April 20-21, 1956)

1. On Certain Stabilities of Sample Survey Response, (Preliminary Report), DAVID ROSENBLATT, American University, (By Title).

In economic and demographic surveys, studies of reporting behavior are sometimes undertaken through reinterviews of identical respondents using a similar or more intense mode of inquiry. Stability of response is examined in the light of cross-tabulation data on first vs. second response. Assume: (1) there exist response observables $\alpha_1, \dots, \alpha_r$, cryptostates β_1, \dots, β_L , generally unobservable, which may, but need not, correspond to response observables; (2) $L \times r$ stochastic matrices Φ_1, Φ_2 , respectively, giving for each