CONTRIBUTIONS TO THE THEORY OF RANK ORDER STATISTICS: COMPUTATION RULES FOR PROBABILITIES OF RANK ORDERS[1]

By I. Richard Savage

University of Minnesota

1. Introduction. For most sampling situations the computation of the non-null probabilities of rank orders involves either difficult multiple integrations or extensive Monte Carlo sampling ([1], [2], [3]). In this note back-recursive rules are given for computing the probabilities of rank orders for the one and two sample problems ([2] (Section 1), [1] (Section 2)). For the one-sample problem the rule permits the computations for samples of size n from the results with samples of $n + 1$. For the two-sample problem the rule permits the computations for samples of size m and n from the results with samples of $m + 1$ and n (m and $n + 1$). Since most computations done analytically are built up from smaller to larger sample sizes these results will, for that case, have limited value, e.g., in checking numerical work. For Monte Carlo sampling, however, there is no reason for starting with the smaller samples and in this case the rules will be of service.

2. One-sample rule. Let $P_n(z)$ be the probability of the rank order $z = (z_1, \cdots, z_n)$, where $z_i = 0(1)$ if the ith smallest of the observed absolute values was from a negative (positive) observed deviation from a hypothetical median, e.g., if the observed deviations are $(2.2, -.7, .5, -1.1, 3.0)$, then $z = (10011)$.

Rule I. To compute $P_n(z)$ add all $[2(n + 1)$ in number$]$ the $P_{n+1}(z^{ij})$ and divide by $(n + 1)$, where

$$z^{ij} = (z_1, \cdots, i, z_j, \cdots, z_n), \qquad i = 0, 1 \text{ and } j = 1, \cdots, n + 1.$$

Note a. Several of the z^{ij} will be the same.

Note b. The rule can be obtained using the analytic expressions for $P_n(z)$ given in [2] (equation 1.1). Another proof can be obtained by noting that, after the sample of size n is formed, an additional observation must fall either between existing observations or before them or after them.

Example I. Numerical results for the one-sample problem are not available. The following, however, suggests the kind of computing formulas that could be used. For $n = 3$,

$$P_3(010) = [P_4(1010) + P_4(0010) + P_4(0110) + P_4(0010) + P_4(0110) + P_4(0100) + P_4(0101) + P_4(0100)]/4.$$

3. Two-sample rule. Let $P_{m,n}(z)$ be the probability of the rank order $z = (z_1, \cdots, z_{m+n})$ where $z_i = 0(1)$ if the ith smallest of the observed values

Received August 28, 1959.

[1] Work done in part under contract Nonr 2582(00), Task NR 042-200 of the Office of Naval Research.

was from the first (second) sample, e.g., if the observed values in the first sample were $(-1.5, 2.6)$ and in the second sample $(3.4, -.9)$, then $z = (0101)$.

RULE II. To compute $P_{m,n}(z)$ add all $[(m + n + 1)$ in number] of the $P_{m+1,n}(z^j)$ and divide by $(m + 1)$, where

$$z^j = (z_1, \cdots, 0, z_j, \cdots, z_{m+n}), j = 1, \cdots, (m + n + 1).$$

Note a. Several of the z^j will be the same.

Note b. The roles of m and n can be interchanged in the obvious manner.

Note c. The rule can be obtained using the analytic expression

$$P_{m,n}(z) = m!n! \int \cdots \int_{-\infty < w_1 \cdots w_{m+n} < \infty} \prod_{i=1}^{m+n} [f^{1-z_i}(w_i) g^{z_i}(w_i)\, dw_i],$$

where $f(w)[g(w)]$ is the density of the first [second] population. Another proof can be obtained by noting that, after the samples of size m and n have been obtained, an additional observation from the first population must either be between a pair of the observations of the original $m + n$ or before or after them.

EXAMPLE II. For the two-sample problem with $m = 3$ and $n = 2$,

$$P_{3,2}(00011) = [P_{3,3}(100011) + P_{3,3}(010011) + P_{3,3}(001011) + 3P_{3,3}(000111)]/3.$$

Teichroew [3] gives .0394 as the exact value, and .0410 as the Monte Carlo value (2000 samples) when the two populations are normal with means differing by $\frac{1}{2}$ of the common standard deviation. Using Teichroew's [3] Monte Carlo results for $m = 3$, $n = 3$ (4000 samples) in the above formula, one obtains $P_{3,2}(00011) = [.03250 + .01825 + .011875 + 3(.01675)]/3 = .03992$. Additional results for $m = 3$, $n = 2$ could be obtained from $m = 4$, $n = 2$ and from $m = 4$, $n = 3$ via $m = 3$, $n = 3$ [3].

REFERENCES

[1] I. RICHARD SAVAGE, "Contributions to the theory of rank order statistics—the two-sample case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 590–615.

[2] I. RICHARD SAVAGE, "Contributions to the theory of rank order statistics—the one-sample case," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 1018–1023.

[3] D. TEICHROEW, "Empirical power functions for nonparametric two-sample tests for small samples," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 340–344.

AN INEQUALITY FOR BALANCED INCOMPLETE BLOCK DESIGNS

BY WADIE F. MIKHAIL

University of North Carolina

1. Summary. The inequality $b \geqq v + r - 1$ for a balanced incomplete block design was proved by Bose [1] under the assumption of resolvability. In this note

Received August 5, 1959; revised October 17, 1959.