SOME RESULTS ON RESTRICTED OCCUPANCY THEORY[1]

By John E. Freund and Arthur N. Pozner

Virginia Polytechnic Institute

1. Introduction. If k judges rate a product on a discrete scale, say 0, 1, 2, $\cdots$, or m, it is not only important to know the average rating assigned to the product, but it is also important to know the consistency of the ratings. Whereas the average rating is r/k, where r is the total number of points assigned to the product by the k judges, it seems reasonable to use the variance of the ratings, i.e.,

$$s^2 = \sum_{i=0}^{m} \left(i - \frac{r}{k}\right)^2 \left(\frac{v_i}{k}\right), \tag{1}$$

as a measure of consistency. Here v_i stands for the number of judges who gave the product the rating i.

In order to test the consistency of such ratings, it will be necessary to find a suitable mathematical model which will assign low probabilities to cases in which the ratings are either very inconsistent or overly consistent. It is felt that the appropriate model is provided by that of restricted occupancy theory, in which we consider as equally likely all possible distributions of r indistinguishable objects among k cells with at most m objects per cell. With this model we shall then test whether it is reasonable to suppose that the r points given by the k judges are randomly distributed among the k judges.

Related problems dealing with the probability that x cells contain more than q objects when r objects are distributed among k cells, and the probaility that x cells contain m objects when r objects are distributed among k cells with at most m objects per cell were investigated by Baticle, [1], [2], and [3], with reference to applications to casualty insurance, merchandizing, and transportation.

2. A restricted occupancy distribution. Restricted occupancy theory deals with the distribution of r objects among k cells if a maximum of m objects is permitted per cell, the cells are distinguishable, and empty cells are permitted. In this paper we shall investigate the distribution of the variables v_i ($i = 0, 1, 2, \cdots, m$), standing for the number of cells occupied by i objects, respectively, if r indistinguishable objects are distributed among k cells with a maximum of m objects per cell.

Since r and k are assumed to be fixed, it should be noted that the variables v_i are subject to the two linear restrictions

$$\sum_{i=0}^{m} v_i = k \quad \text{and} \quad \sum_{i=0}^{m} i \cdot v_i = r, \tag{2}$$

and any $m - 1$ of the v_i will thus determine the remaining two.

Received December 30, 1954, revised November 16, 1955.

[1] Research sponsored by the Office of Ordnance, U.S. Army, Contract No. DA-36-034-ORD-1527RD.

If we now write as $N(r, k, m)$ the total number of ways in which r indistinguishable objects can be distributed among k cells with at most m objects per cell, we find that this *restricted occupancy coefficient* is the coefficient of x^r in

$$f(x; k, m) = (1 + x + x^2 + \cdots + x^m)^k. \tag{3}$$

Since

$$\begin{aligned} f(x; k, m) &= (1 - x^{m+1})^k(1 - x)^{-k} \\ &= (1 - x^{m+1})^k \cdot \sum_{n=0}^{\infty} \binom{-k}{n} (-1)^n x^n \\ &= (1 - x^{m+1})^k \cdot \sum_{n=0}^{\infty} \binom{k + n - 1}{k - 1} x^n, \end{aligned}$$

we can write explicitly

$$N(r, k, m) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \binom{k + r - j(m + 1) - 1}{k - 1}. \tag{4}$$

If we now let $i_1, i_2, \cdots, i_{m-1}$ stand for an arbitrary permutation of any $m - 1$ of the $m + 1$ numbers $0, 1, 2, \cdots, m$, the joint distribution of the v_{i_j}, for $j = 1, 2, \cdots, m - 1$, may be written as

$$f(v_{i_1}, \cdots, v_{i_{m-1}}) = \frac{\binom{k}{v_{i_1}, v_{i_2}, \cdots, v_{i_{m-1}}} \cdot \binom{k - \sum_{j=1}^{m-1} v_{i_j}}{q}}{N(r, k, m)}, \tag{5}$$

where

$$q = \frac{k \cdot i_{m+1} - r - \sum_{j=1}^{m-1} (i_{m+1} - i_j) v_{i_j}}{i_{m+1} - i_m},$$

and

$$\binom{k}{v_{i_1}, v_{i_2}, \cdots, v_{i_{m-1}}}$$

is a multinomial coefficient. Also, i_m and i_{m+1} are the two numbers which were omitted when choosing the $m - 1$ numbers i_j.

The moments of the v_{i_j} may be obtained directly from (5), but their derivation is simplified if we refer to the generating function given in (3). Writing

$$\begin{aligned} f(x; k, m) &= [x^i + (1 + x + x^2 + \cdots + x^m - x^i)]^k \\ &= \sum_{j=0}^{k} \binom{k}{j} x^{ij}(1 + x + x^2 + \cdots + x^m - x^i)^{k-j}, \end{aligned}$$

we find that for a fixed value of j, the coefficient of x^r represents the number of ways in which r indistinguishable objects can be distributed among k cells with at most m objects per cell *under the condition that exactly j cells have exactly i occupants*. Writing this coefficient as $N(r, k, m|v_i = j)$, the marginal distribution of v_i becomes

$$\text{(6)} \qquad \Pr\{v_i = j\} = \frac{N(r, k, m \mid v_i = j)}{N(r, k, m)}.$$

If we let $E[v_i^{(t)}]$ stand for the tth factorial moment of v_i, we have

$$\text{(7)} \qquad N(r, k, m)E[v_i^{(t)}] = \sum_{j=0}^{k} j^{(t)}N(r, k, m) \mid v_i = j),$$

and the right-hand member of (7) is the coefficient of x^r in

$$\sum_{j=0}^{k} \binom{k}{j} j^{(t)} x^{ij}(1 + x + x^2 + \cdots + x^m - x^i)^{k-j} = k^{(t)}(1 + x + x^2 + \cdots + x)^{k-t}x^{it},$$

where $j^{(t)}$ and $k^{(t)}$ are the tth factorial powers of j and k. We thus find that

$$\text{(8)} \qquad E[v_i^{(t)}] = \frac{k^{(t)}N(r - ti, k - t, m)}{N(r, k, m)}.$$

Similarly, if we let $N(r, k, m|v_{i_1} = j_1, v_{i_2} = j_2)$ stand for the number of ways in which r indistinguishable objects can be distributed among k cells with at most m objects per cell *under the condition that exactly j_1 cells have exactly i_1 occupants and exactly j_2 cells have exactly i_2 occupants* (with i_1 not equal to i_2), this quantity is given by the coefficient of x^r in the corresponding term of

$$\begin{aligned} f(x; k, m) &= [x^{i_1} + x^{i_2} + (1 + x + \cdots + x^m - x^{i_1} - x^{i_2})]^k \\ &= \sum_{j_1=0}^{k} \sum_{j_2=0}^{k} \binom{k}{j_1, j_2} x^{i_1 j_1 + i_2 j_2}(1 + x + \cdots + x^m - x^{i_1} - x^{i_2})^{k-j_1-j_2}. \end{aligned}$$

Using the same steps as above, it may then be shown that

$$\text{(9)} \qquad E[v_{i_1}^{(t_1)} v_{i_2}^{(t_2)}] = \frac{k^{(t_1+t_2)}N(r - i_1t_1 - i_2t_2, k - t_1 - t_2, m)}{N(r, k, m)},$$

and higher factorial product moments can be obtained in the same way.

The calculation of the restricted-occupancy coefficients needed in the evaluation of the moments of the v_{i_j} and subsequently of the mean and variance of s^2 [as defined in (1)] is greatly simplied by the use of the recursion relations

$$\text{(10a)} \qquad N(r, k, m) = \sum_{i=0}^{m} N(r - i, k - 1, m),$$

$$\text{(10b)} \qquad N(r, k - 1, m) = \sum_{i=1}^{m} \frac{ki - r}{r} \cdot N(r - i, k - 1, m),$$

or

$$\text{(10c)} \qquad N(r, k, m) = N(r - 1, k, m) + N(r, k - 1, m) - N(r - m - 1, k - 1, m).$$

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

$$f(x; k, m) = f(x; 1, m)f(x; k-1, m), \tag{11a}$$

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

or

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

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 THEOREM[1]

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 $\|C_{n1}(i, j)\|, \cdots, \|C_{nk}(i, j)\|$ be $n \times n$ matrices of real numbers. Also let $(R_1, \cdots, R_n)$ be a random variable which takes each permutation of $(1, \cdots, n)$ 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.