STATISTICAL TESTS BASED ON PERMUTATIONS OF THE OBSERVATIONS

A. WALD AND J. WOLFOWITZ

Columbia University

1. Introduction. One of the problems of statistical inference is to devise exact tests of significance when the form of the underlying probability distribution is unknown. The idea of a general method of dealing with this problem originated with R. A. Fisher [13, 14]. The essential feature of this method is that a certain set of permutations of the observations is considered, having the property that each permutation is equally likely under the hypothesis to be tested. Thus, an exact test on the level of significance α can be constructed by choosing a proportion α of the permutations as critical region. In an interesting paper H. Scheffé [2] has shown that for a general class of problems this is the only possible method of constructing exact tests of significance.

Tests based on permutations of the observations have been proposed and studied by R. A. Fisher, E. J. G. Pitman, B. L. Welch, the present authors, and others. Pitman and Welch derived the first few moments of the statistics used in their test procedures. However, it is desirable to derive at least the limiting distributions of these statistics and make it practicable to carry out tests of significance when the sample is large. Such a large sample distribution was derived for a statistic considered elsewhere [1] by the present authors.

In this paper a general theorem on the limiting distribution of linear forms in the universe of permutations of the observations is derived. As an application of this theorem, the limiting distributions of the rank correlation coefficient and that of several statistics considered by Pitman and Welch, are obtained. In the last section the limiting distribution of Hotelling's generalized T in the universe of permutations of the observations is derived.

2. A theorem on linear forms. Let $H_N = (h_1, h_2 \dots, h_N)$ $(N = 1, 2, \dots,$ ad inf.) be sequences of real numbers and let

$$\mu_r(H_N) = N^{-1} \sum_{\alpha=1}^{N} \left((h_\alpha - N^{-1} \sum_{\beta=1}^{N} h_\beta \right)^r$$

for all integral values of r. We define the following symbols in the customary manner: For any function $f(N)$ and any positive function $\varphi(N)$ let $f(N) = O(\varphi(N))$ mean that $| f(N)/\varphi(N) |$ is bounded from above for all N and let

$$f(N) = \Omega(\varphi(N))$$

mean that

$$f(N) = O(\varphi(N))$$

and that

$$\lim_N \inf | f(N)/\varphi(N) | > 0.$$

Also let

$$f(N) = o(\varphi(N))$$

mean that

$$\lim_{N \to \infty} f(N)/\varphi(N) = 0.$$

Let $[\rho]$ denote the largest integer $\leq \rho$.

We shall say that the sequences $H_N(N = 1, 2, \cdots$, ad inf.) satisfy the condition W if, for all integral $r > 2$,

$$\frac{\mu_r(H_N)}{[\mu_2(H_N)]^{r/2}} = O(1). \tag{2.1}$$

For any value of N let

$$X = (x_1, x_2, \cdots, x_N)$$

be a chance variable whose domain of definition is made up of the $N!$ permutations of the elements of the sequence $A_N = (a_1, a_2, \cdots, a_N)$. (If two of the $a_i(i = 1, 2, \cdots, N)$ are identical we assume that some distinguishing index is attached to each so that they can then be regarded as distinct and so that there still are $N!$ permutations of the elements $a_1, \cdots, a_N$). Let each permutation of A_N have the same probability $(N!)^{-1}$. Let $E(Y)$ and $\sigma^2(Y)$ denote, respectively, the expectation and variance of any chance variable Y.

We now prove the following:

THEOREM. *Let the sequences* $A_N = (a_1, a_2, \cdots, a_N)$ *and* $D_N = (d_1, d_2, \cdots, d_N)$ ($N = 1, 2, \cdots$, ad inf.) *satisfy the condition* W. *Let the chance variable* L_N *be defined as*

$$L_N = \sum_{i=1}^{N} d_i x_i .$$

Then as $N \to \infty$, *the probability of the inequality*

$$L_N - E(L_N) < t\,\sigma\,(L_N)$$

for any real t, approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}x^2}\, dx.$$

For convenience the proof will be divided into several lemmas.

Since

$$L_N^* = \frac{L_N - E(L_N)}{\sigma(L_N)}$$

remains invariant if a constant is added to all the elements of D_N or of A_N, or if the elements of either of the latter are multiplied by any constant other than zero, we may, in the formation of L_N^*, replace A_N and D_N by the sequences A_N'

and D'_N, respectively, whose ith elements a'_i and $d'_i (i = 1, 2, \cdots, N)$ are, respectively

$$a'_i = [\mu_2(A_n)]^{-\frac{1}{2}}\left(a_i - N^{-1}\sum_{j=1}^{N} a_j\right) \tag{2.2}$$

and

$$d'_i = [\mu_2(D_N)]^{-\frac{1}{2}}\left(d_i - N^{-1}\sum_{j=1}^{N} d_j\right). \tag{2.3}$$

The sequences A'_N and D'_N satisfy the condition W. Furthermore,

$$\mu_1(A'_N) \equiv \mu_1(D'_N) \equiv 0 \tag{2.4}$$

and

$$\mu_2(A'_N) \equiv \mu_2(D'_N) \equiv 1. \tag{2.5}$$

LEMMA 1.

$$\sum_{\alpha_1<\alpha_2<\cdots<\alpha_k\le N}\cdots\sum a'_{\alpha_1}a'_{\alpha_2}\cdots a'_{\alpha_k} = O(N^{[k/2]}) \tag{2.6}$$

$$\sum_{\alpha_1<\alpha_2<\cdots<\alpha_k\le N}\cdots\sum d'_{\alpha_1}d'_{\alpha_2}\cdots d'_{\alpha_k} = O(N^{[k/2]}). \tag{2.7}$$

From (2.4), (2.5), and the fact that the A'_N and D'_N satisfy condition W, it follows that the A'_N and D'_N satisfy conditions a), b), and c) of the theorem on page 383 of [1]. Our lemma 1 is the same as lemma 1 of [1].

LEMMA 2. *Let*

$$V = (v_1, v_2, \cdots, v_N)$$

be the same permutation of the elements of A'_N *that* X *is of the elements of* A_N. *Let* $y = v_1 \cdots v_k z$ *where* $z = v_{(k+1)}^{i_1} \cdots v_{(k+r)}^{i_r}$, $i_j > 1$ $(j = 1, 2, \cdots, r)$, *and* $k, r, i_1, \cdots, i_r$ *are fixed values independent of* N.
Then

$$E(y) = O(N^{[k/2]-k}). \tag{2.8}$$

This is Lemma 2 of [1].

In a similar manner we obtain that

$$\sum_{\alpha_1,\alpha_2,\cdots,\alpha_{(k+r)}}\cdots\sum d'_{\alpha_1}\cdots d'_{\alpha_k}d'^{i_1}_{\alpha_{(k+1)}}\cdots d'^{i_r}_{\alpha_{(k+r)}} = O(N^{[k/2]-k})\cdot O(N^{k+r}) = O(N^{[k/2]+r}). \tag{2.9}$$

The summation in the above formula is to be taken over all possible sets of $k + r$ distinct positive integers $\le N$.

LEMMA 3. *Let* $\alpha_1, \cdots, \alpha_{(k+r)}$ *be* $(k + r)$ *distinct positive integers* $\le N$. *Then*

$$E(v_1v_2\cdots v_k v_{(k+1)}^{i_1}\cdots v_{(k+r)}^{i_r}) = E(v_{\alpha_1}v_{\alpha_2}\cdots v_{\alpha_k}v_{\alpha_{(k+1)}}^{i_1}\cdots v_{\alpha_{(k+r)}}^{i_r}). \tag{2.10}$$

This follows from the fact that all permutations of A'_N have the same probability.

LEMMA 4. *Let*

$$L'_N = \sum_{i=1}^{N} d'_i v_i .$$

Then

$$E(L'^p_N) = O(N)^{[p/2]}). \tag{2.11}$$

PROOF: Expand L'^p_N and take the expected value of the individual terms. The contribution to $E(L'^p_N)$ of all the terms which are multiples of the type appearing in the right member of (2.10) with fixed $k, r, i_1, \cdots, i_r (k + i_1 + \cdots + i_r = p)$, is, by Lemmas 2 and 3

$$O(N^{[k/2]-k}) \cdot \sum_{\substack{\alpha_1, \cdots, \alpha_{(k+r)} \\ \text{all different}}} \cdots \sum d'_{\alpha_1} \cdots d'_{\alpha_k} d'^{i_1}_{\alpha(k+1)} \cdots d'^{i_r}_{\alpha(k+r)} = O(N^{[k/2]-k}) O(N^{[k/2]+r})$$

$$= O(N^{2[k/2]-k+r}).$$

Since $i_j > 1 (j = 1, \cdots, r)$, it follows from the fact that $k + i_1 + \cdots + i_r = p$ that $2r \leq p - k$ and that $2r = p - k$ only if $i_1 = \cdots = i_r = 2$. Now

$$2\left[\frac{k}{2}\right] - k + r \leq r \leq \frac{p-k}{2} \leq \frac{p}{2}. \tag{2.12}$$

Hence the maximum value of $2\left[\frac{k}{2}\right] - k + r$ is reached when $r = \left[\frac{p}{2}\right]$ and $k = 0$. This proves the lemma.

From the last remarks of the preceding paragraph we obtain

LEMMA 5.

$$E(L'^{2j}_N) - \frac{(2j)!}{j!2^j} (\sum_{\substack{\alpha_1, \cdots, \alpha_j \\ \text{all different}}} \cdots \sum d'^2_{\alpha_1} \cdots d'^2_{\alpha_j}) E(v_1^2 \cdots v_j^2) = o(N^j). \tag{2.13}$$

We now prove

LEMMA 6.

$$E(L'_N) = 0 \tag{2.14}$$

$$E(L'^2_N) = NE(v_1^2) + o(N) = N + o(N). \tag{2.15}$$

Equation (2.14) follows from (2.2). Consider the expectations of the various terms in the expansion of L'^2_N. The sum of all the terms of the type

$$d'_i d'_j E(v_i v_j)$$

is

$$(\sum_{i \neq j} d'_i d'_j) E(v_1 v_2) = O(N) O(N^{-1}) = O(1),$$

by Lemmas 1 and 2. The sum of all the terms of the type

$$d'^2_i E(v_i^2)$$

is

$$\left(\sum_{i=1}^{N} d_i'^2\right) E(v_1^2) = NE(v_1^2) = N,$$

by (2.2) and (2.3). This proves the lemma.

LEMMA 7.

$$E(v_1^2 \cdots v_j^2) = 1 + o\,(1) \tag{2.16}$$

$$\sum_{\substack{\alpha_1 \cdots \alpha_j \\ \text{all different}}} \cdots \sum d_{\alpha_1}'^2 \cdots d_{\alpha_j}'^2 = N^j + o(N^j). \tag{2.17}$$

From (2.2) and (2.3), and Lemma 3, it follows that it will be sufficient to prove (2.17), because (2.16) follows in the same manner. Consider the relation

$$N^j = \left(\sum_{i=1}^{N} d_i'^2\right)^j = \sum_{\substack{\alpha_1, \cdots, \alpha_j \\ \text{all different}}} \cdots \sum d_{\alpha_1}'^2 \cdots d_{\alpha_j}'^2 + \text{other terms.}$$

By (2.9) the sum of these other terms must be not larger than $O(N^{j-1})$. From this follows the lemma.

PROOF of the theorem: Since

$$L_N^* = \frac{L_N'}{\sigma(L_N')} = \frac{L_N - E(L_N)}{\sigma(L_N)},$$

it will be sufficient to show that the moments of L_N^* approach those of the normal distribution as $N \to \infty$. From (2.14), (2.15), and (2.11) we see that, when p is odd, the pth moment of L_N^* is $O(N^{-\frac{1}{2}})$ and hence approaches zero as $N \to \infty$. When p is even and $= 2s$ (say), it follows from Lemma 5 that

$$E(L_N'^{2s}) - \frac{(2s)!}{s!2^s} \Big(\sum_{\substack{\alpha_1, \cdots, \alpha_s \\ \text{all different}}} \cdots \sum d_{\alpha_1}'^2 \cdots d_{\alpha_s}'^2\Big) E(v_1^2 \cdots v_s^2) = o(N^s).$$

Hence from (2.16) and (2.17)

$$E(L_N'^{2s}) = \frac{(2s)!}{s!2^s} N^s + o(N^s). \tag{2.18}$$

From (2.18) and (2.15) we obtain that

$$\lim_{N \to \infty} E(L_N^{*2s}) = \frac{(2s)!}{s!2^s}.$$

This completes the proof of the theorem.

It will be noticed that nothing in the foregoing proof requires that, when $N < N'$, the sequences A_N and D_N be subsequences of $A_{N'}$ and $D_{N'}$. Indeed, the sequences were written as they were simply for typographic brevity. We have therefore

COROLLARY 1. *The theorem is valid for sequences*

$$A_N = (a_{N1}, \cdots, a_{NN})$$
$$D_N = (d_{N1}, \cdots, d_{NN})$$
$$(N = 1, 2, \cdots \text{ ad inf.})$$

provided they fulfill condition W.

COROLLARY 2. *If the elements* $a_i (i = 1, 2, \cdots$ ad inf.) *are all independent observations on the same chance variable, all of whose moments are finite and whose variance is positive, the sequences* $A_N(N = 1, 2, \cdots$, ad inf.) *will fulfill condition W with probability one.*

3. The rank correlation coefficient. For this well known statistic (see [3])

$$A_N \equiv D_N \equiv (1, 2, 3, \cdots, N).$$

The sequences A_N and D_N satisfy the condition W. For

$$\sum_{i=1}^{N} i^r = O(N^{r+1})$$

and hence, for $r \geq 3$

$$\mu_r(A_N) = \mu_r(D_N) = O(N^r).$$

Also

$$\mu_2(A_N) = \mu_2(D_N) = \Omega(N^2).$$

Hence the distribution of the rank correlation coefficient is asymptotically normal in the case of statistical independence. This result was first proved by Hotelling and Pabst [3].

4. Pitman's test for dependence between two variates. The distribution of the correlation coefficient in the population of permutations of the observations was used by Pitman [4] in a test for dependence between two variates which "involves no assumptions" about the distributions of these variates. In our notation, let

$$(a_i, d_i)(i = 1, 2, \cdots, N)$$

be N observations on the pair of variates A and D whose dependence it is desired to test. Then the value of the correlation coefficient is

$$N^{-1} \sum_{i=1}^{N} d_i' a_i' .$$

At the level β the observations are considered to be significant if the probability that $N^{-1} | L_N' |$ be equal to or greater than the absolute value of the actually observed correlation coefficient is $\leq \beta$.

In his paper ([4], page 227] Pitman points out that if the ratios of certain sample cumulants are "not too large," then, as $N \to \infty$, the first four moments of $N^{-\frac{1}{2}} L'_N$ will approach 0, 1, 0, and 3, respectively (the first moment is always zero). Our theorem and the relation (2.15) make clear that under proper circumstances all the moments will approach those of the normal distribution.

5. Pitman's procedure for testing the hypothesis that two samples are from the same population. For testing the hypothesis that two samples came from the same population Pitman [5] proposed the following procedure:
Let one sample be

$$a_1, a_2, \cdots, a_m$$

and the other

$$a_{m+1}, a_{m+2}, \cdots, a_{m+n}.$$

Write $m + n = N$, and construct the sequences A_N and A'_N as before defined. Let

$$d_i = 1 \qquad (i = 1, \cdots, m)$$

$$d_i = 0 \qquad (i = m + 1, \cdots, N)$$

and construct the sequences D_N and D'_N. Then the value of the statistic considered by Pitman is, except for a constant factor,

$$N^{-\frac{1}{2}} \left(\sum_{i=1}^{N} d'_i a'_i \right). \tag{5.1}$$

At the level β the observations are considered significant if the probability that $N^{-\frac{1}{2}} \mid L'_N \mid$ be equal to or greater than the observed absolute value of the expression (5.1) is $\leq \beta$.

Let $N \to \infty$, while $\frac{m}{n}$ is constant. Then the sequences D_N are seen to satisfy condition W. If then the sequences A_N satisfy condition W we may, for large N, employ the result of our theorem and expeditiously determine the critical value of Pitman's statistic.

6. Analysis of variance in randomized blocks. Welch [7] and Pitman [6] consider the following problem: Each of n different "varieties of a plant" is planted in one of the n cells which constitute a "block." It is desired to test, on the basis of results from m blocks, the null hypothesis that there is no difference among the varieties. In order to eliminate a possible bias caused by variations in fertility among the cells of a block, the varieties are assigned at random to the cells of a block. If the cells of the jth block are designated by $(j1), (j2), \cdots, (jn)$, a permutation of the integers $1, 2, \cdots, n$ is allocated to the jth block by a chance process, each permutation having the same probability $(n!)^{-1}$.

Let x_{ijk} be the yield of the ith variety in the kth cell of the jth block to which it was assigned by the randomization process. It is assumed that

$$x_{ijk} = y_{jk} + \delta_i + \epsilon_{jk},$$

where y_{jk} is the "effect" of the kth cell in the jth block, δ_i is the "effect" of the ith variety, and ϵ_{jk} are chance variables about whose distribution we assume nothing. The null hypothesis states that

$$\delta_1 = \delta_2 = \cdots = \delta_n = 0.$$

Let a_{jk} be the yield in the kth cell of the jth block and x_{ij} the yield of the ith variety in the jth block. If the null hypothesis is true then, because of the randomization within each block described above, the conditional probability that, given the set $\{a_{jk}\}$ $(k = 1, 2, \cdots, n)$, the sequence $x_{1j}, x_{2j} \cdots, x_{nj}$, be any given permutation of the elements of $\{a_{jk}\}$ is $(n!)^{-1}$. Permuting in all the blocks simultaneously we have that, under the null hypothesis, given the set of mn values $\{a_{jk}\}$ $(j = 1, 2, \cdots m;\ k = 1, 2, \cdots, n)$, the conditional probability of any of the permutations is the same, $(n!)^{-m}$. This permits an exact test of the null hypothesis.

The classical analysis of variance statistic that would be employed in the conventional two-way classification with independent normally distributed observations is

$$F = \frac{(m-1)m \sum (x_{i\cdot} - x)^2}{\sum \sum (x_{ij} - x_{i\cdot} - x_{\cdot j} + x)^2}$$

where

$$x_{i\cdot} = m^{-1} \sum_j x_{ij} \qquad x = (mn)^{-1} \sum \sum x_{ij}.$$
$$x_{\cdot j} = n^{-1} \sum_i x_{ij}$$

The statistic W used by Welch and Pitman is

$$W = F(m - 1 + F)^{-1}.$$

Since W is a monotonic function of F and the critical regions are the upper tails, the two tests are equivalent. The distribution of F or W is to be determined in the same manner as that of the other statistics discussed in this paper, i.e., over the equally probable permutations of the values actually observed. The critical region is, as usual, the upper tail.

Since x_{ij} takes any of the values $a_{j1}, \cdots, a_{jn}$ with probability $1/n$, we have

$$E(x_{ij}) = n^{-1} \sum_k a_{jk} = a_j \quad \text{(say)}. \tag{6.1}$$

$$\sigma^2(x_{ij}) = n^{-1} \sum_k (a_{jk} - a_j)^2 = b_j \quad \text{(say)}. \tag{6.2}$$

$$\begin{aligned} \sigma(x_{i_1j} x_{i_2j}) &= [n(n-1)]^{-1} \sum_{k_1 \neq k_2} a_{jk_1} a_{jk_2} - a_j^2 \\ &= [n(n-1)]^{-1} [(\sum_k a_{jk})^2 - \sum_k a_{jk}^2] - a_j^2 \\ &= [n^2 a_j^2 - \sum_k a_{jk}^2][n(n-1)]^{-1} - a_j^2 \\ &= (n-1)^{-1}[a_j^2 - n^{-1} \sum_k a_{jk}^2] = -(n-1)^{-1} b_j. \end{aligned} \tag{6.3}$$

Hence

$$E(x_{i\cdot}) = m^{-1} \sum a_j . \tag{6.4}$$

$$\sigma^2(x_{i\cdot}) = m^{-2} \sum b_j = b \quad \text{(say)}. \tag{6.5}$$

$$\sigma(x_{i_1\cdot} x_{i_2\cdot}) = -[m^2(n-1)]^{-1} \sum b_j = c \quad \text{(say)}. \tag{6.6}$$
$$i_1 \neq i_2$$

Let

$$x_{ij}^* = \sum_v \lambda_{iv} x_{vj} \qquad (i, v = 1, \cdots, n)$$

where $|| \lambda_{iv} ||$ is an orthogonal matrix and

$$\lambda_{n1} = \lambda_{n2} = \cdots = \lambda_{nn} = n^{-\frac{1}{2}}.$$

Then it follows that

$$\begin{aligned} E(x_{i\cdot}^*) &= 0 \\ \sigma^2(x_{i\cdot}^*) &= b - c \qquad (i = 1, 2, \cdots, n-1) \\ \sigma(x_{i_1\cdot}^* x_{i_2\cdot}^*) &= 0 \qquad (i_1 \neq i_2 ; i_1, i_2 = 1, \cdots, n-1). \end{aligned} \tag{6.7}$$

Furthermore, we have

$$\sum_{i=1}^{n-1} x_{i\cdot}^{*2} = \sum_{i=1}^{n} (x_{i\cdot} - x)^2. \tag{6.8}$$

Applying the well known identity

$$\Sigma\Sigma(x_{ij} - x_{i\cdot} - x_{\cdot j} + x)^2 = \Sigma\Sigma(x_{ij} - x_{\cdot j})^2 - m\Sigma(x_{i\cdot} - x)^2$$

to the definitions of F and W we obtain

$$W = \frac{m \sum_i (x_{i\cdot} - x)^2}{\sum_i \sum_j (x_{ij} - x_{\cdot j})^2}. \tag{6.9}$$

The denominator of the right member of (6.9) is invariant under permutations *within* each block and equals

$$\sum_j \sum_k (a_{jk} - a_j)^2 = (n-1)m^2(b-c).$$

Hence

$$\begin{aligned} W &= [m(n-1)(b-c)]^{-1} \sum_{i=1}^{n} (x_{i\cdot} - x)^2 \\ &= [m(n-1)(b-c)]^{-1} \sum_{i=1}^{n-1} x_{i\cdot}^{*2}. \end{aligned} \tag{6.10}$$

If the joint distribution of the $x_{i\cdot}^*$ $(i = 1, 2, \cdots, n-1)$ over the set of admissible permutations approaches a normal distribution with non-singular correlation matrix as m, the number of blocks, becomes large, it follows from (6.7) and (6.10) that the distribution of $m(n-1)W$ approaches the x^2 distribution with $n-1$ degrees of freedom. Hence it remains to indicate conditions on the set $\{a_{jk}\}$ which would make the distribution of the x_i^* approach normality. Each $x_{i.}^*$ is the mean of independent variables, so these conditions need not be very restrictive.

According to Cramér [8], Theorem 21a, page 113, if the variances and covariances fulfill certain requirements (the limiting correlation matrix should also be non-singular) and if a generalized Lindeberg condition holds, normality in the limit will follow. Somewhat more restrictive conditions which are simpler to state and which will be satisfied in most statistical applications are that $o < c' < b_j < c''$ for all j, where c' and c'' are positive constants. Since the variance of x_{ij}^* is $(n-1)^{-1}nb_j$, it can be seen that the above inequalities imply the fulfillment of the conditions of the Laplace-Liapounoff theorem (see, for example, Uspensky [9], page 318). By [6.7] the correlation matrix is non-singular.

7. Hotelling's generalized T for permutation of the observations. In this section we shall restrict ourselves to bivariate populations, the extension to more than two variables being straightforward. Let $(u_{11}, u_{21}), \cdots, (u_{1m}, u_{2m})$ be m pairs of observations on the chance variables U_1, U_2, and $(u_{1(m+1)}, u_{2(m+1)}), \cdots, (u_{1N}, u_{2N})$, be n pairs of observations on the chance variables U_1', U_2', where $m + n = N$. If each of the pairs U_1, U_2, and U_1', U_2' is jointly normally distributed with the same convariance matrix, the Hotelling generalized T for testing the null hypothesis that

$$E(U_1) = E(U_1') \tag{7.1}$$

and

$$E(U_2) = E(U_2'), \tag{7.2}$$

is given (Hotelling [10]) by

$$T^2 = N^{-1}(mn) \sum_{j=1}^{2} \sum_{i=1}^{2} q_{ij}(\bar{u}_i - \bar{u}_i')(\bar{u}_j - \bar{u}_j')$$

where

$$m\bar{u}_i = \sum_{l=1}^{m} u_{il} \quad n\bar{u}_i' = \sum_{l=m+1}^{N} u_{il}$$

and the matrix $|| q_{ij} ||$ is the inverse of the matrix $|| b_{ij} ||$ with b_{ij} given by

$$(N-2)b_{ij} = \sum_{l=1}^{m} (u_{il} - \bar{u}_i)(u_{jl} - \bar{u}_j) + \sum_{l=m+1}^{N} (u_{il} - \bar{u}_i')(u_{jl} - \bar{u}_j').$$

In Hotelling's procedure the b_{ij} are sample estimates of the population covariances whose distribution is independent of that of the sample means. A constant multiple of the statistic T^2 has the analysis of variance distribution under the null hypothesis. If the population covariances were known and used in place of the b_{ij}, T^2 would have the χ^2 distribution with two degrees of freedom.

Let us now apply the generalized T over the permutations of the actually observed values, as was done with other statistics in previous sections. If we do this literally we will find that the b_{ij} are no longer independent of the sample means. To avoid this complication we shall use a slightly different statistic T' which, as will be shown later, is a monotonic function of T, so that the test based on T' is identical with that based on T. The statistic T' is defined as follows: Let

$$\bar{U}_i = N^{-1} \sum_{k=1}^{N} u_{ik}$$

$$c'_{ij} = N[(N-1)mn]^{-1} \sum_{k=1}^{N} (u_{ik} - \bar{U}_i)(u_{jk} - \bar{U}_j) \qquad (i, j, = 1, 2)$$

and

$$\| q'_{ij} \| = \| c'_{ij} \|^{-1}.$$

Then

$$T'^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} q'_{ij}(\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j). \tag{7.3}$$

The expression T'^2 is much simpler than T^2, since the coefficients q'_{ij} are constants in the population of permutations of the observations. We shall show that T'^2 is a monotonic function of T^2. Let

$$Q_{ij} = \sum_{k=1}^{m} (u_{ik} - \bar{u}_i)(u_{jk} - \bar{u}_j) + \sum_{k=m+1}^{N} (u_{ik} - \bar{u}'_i)(u_{jk} - \bar{u}'_j)$$

$$Q'_{ij} = \sum_{k=1}^{N} (u_{ik} - \bar{U}_i)(u_{jk} - \bar{U}_j)$$

$$\| Q^{ij} \| = \| Q_{ij} \|^{-1}$$

$$\| Q'^{ij} \| = \| Q'_{ij} \|^{-1}.$$

Then the expressions

$$T_1^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} Q^{ij}(\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j) \tag{7.4}$$

and

$$T_2^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} Q'^{ij}(\bar{u}_i - \bar{u}'_i)(\bar{u}_j - \bar{u}'_j), \tag{7.5}$$

are constant multiples of T^2 and T'^2, respectively. Hence it is sufficient to show that T_2^2 is a monotonic function of T_1^2. We have

$$Q'_{ij} = Q_{ij} + m(\bar{u}_i - \bar{U}_i)(\bar{u}_j - \bar{U}_j) + n(\bar{u}'_i - \bar{U}_i)(\bar{u}'_j - \bar{U}_j). \tag{7.6}$$

Furthermore, we have

$$\bar{u}_i - \bar{U}_i = \bar{u}_i - \frac{m\bar{u}_i + n\bar{u}'_i}{m+n} = \frac{n(\bar{u}_i - \bar{u}'_i)}{m+n}. \tag{7.7}$$

Similarly

$$\bar{u}_i' - \bar{U}_i = \bar{u}_i' - \frac{m\bar{u}_i + n\bar{u}_i'}{m + n} = \frac{m(\bar{u}_i' - \bar{u}_i)}{m + n}. \tag{7.8}$$

From (7.6), (7.7) and (7.8) it follows that

$$\begin{aligned} Q'_{ij} &= Q_{ij} + \frac{mn^2}{(m + n)^2}(\bar{u}_i - \bar{u}_i')(\bar{u}_j - \bar{u}_j') + \frac{nm^2}{(m + n)^2}(\bar{u}_i - \bar{u}_i')(\bar{u}_j - \bar{u}_j') \\ &= Q_{ij} + \frac{mn}{m + n}(\bar{u}_i - \bar{u}_i')(\bar{u}_j - \bar{u}_j'). \end{aligned} \tag{7.9}$$

Denote $\frac{mn}{m + n}$ by λ and $\bar{u}_i - \bar{u}_i'$ by h_i. Then we have

$$Q'_{ij} = Q_{ij} + \lambda h_i h_j. \tag{7.10}$$

Denote the cofactor of Q_{ij} in $\| Q_{ij} \|$ by R_{ij} and the cofactor of Q'_{ij} in $\| Q'_{ij} \|$ by R'_{ij}. Then

$$\frac{| Q_{ij} |}{| Q'_{ij} |} = \frac{| Q_{ij} |}{| Q_{ij} + \lambda h_i h_j |} = \frac{| Q_{ij} |}{| Q_{ij} | + \lambda \Sigma\Sigma R_{ij} h_i h_j} = \frac{1}{1 + \lambda T_1^2}. \tag{7.11}$$

Furthermore, we have

$$\frac{| Q_{ij} |}{| Q'_{ij} |} = \frac{| Q'_{ij} - \lambda h_i h_j |}{| Q'_{ij} |} = \frac{| Q'_{ij} | - \lambda \Sigma\Sigma R'_{ij} h_i h_j}{| Q'_{ij} |} = 1 - \lambda T_2^2. \tag{7.12}$$

From (7.11) and (7.12) it follows that T_2^2 is a monotonic function of T_1^2. Hence also T'^2 is a monotonic function of T^2 and, therefore, we do not change our test procedure by using T'^2 instead of T^2.

Let the sequence of pairs

$$(x_{11} . x_{21}), \cdots, (x_{1N}, x_{2N})$$

be a permutation of the actually observed pairs

$$(u_{11}, u_{21}), \cdots, (u_{1N}, u_{2N})$$

where to each permutation is ascribed the same probability $(N!)^{-1}$. Then one obtains for $i = 1, 2$,

$$E(\bar{x}_i - \bar{x}_i') = 0 \tag{7.13}$$

$$\sigma^2(\bar{x}_i - \bar{x}_i') = N[(N - 1)mn]^{-1} \sum_{j=1}^{N} (u_{ij} - \bar{U}_i)^2 = c'_{ii} \tag{7.14}$$

$$E(\bar{x}_1 - \bar{x}_1')(\bar{x}_2 - \bar{x}_2') = N[(N - 1)mn]^{-1} \sum_{j=1}^{N} (u_{1j} - \bar{U}_1)(u_{2j} - \bar{U}_2) = c'_{12}. \tag{7.15}$$

Hence $\| c'_{ij} \|$ is the covariance matrix of the variates

$$(\bar{x}_1 - \bar{x}_1') \qquad \text{and} \qquad (\bar{x}_2 - \bar{x}_2').$$

Now we shall show that the limiting distribution of T'^2, as $N \to \infty$, is the χ^2 distribution with 2 degrees of freedom, provided that the observation u_{ik} $(i = 1, 2; k = 1, \cdots, N)$ satisfy some slight restrictions. Since $\| q'_{ij} \|$ is the inverse of the covariance matrix $\| c'_{ij} \|$ our statement about the limiting distribution of T'^2 is obviously proved if we can show that $\bar{x}_1 - \bar{x}'$ and $\bar{x}_2 - \bar{x}'_2$ have a joint normal distribution in the limit.

Let $N \to \infty$ while m/n remains constant. Let the sequences A_N and D_N of Section II be defined as follows:
There are two sequences A_N, denoted respectively by A_{1N} and A_{2N}, such that

$$a_{ij} = u_{ij} \qquad (i = 1, 2; j = 1, \cdots, N).$$

Also

$$d_j = \frac{1}{m} \qquad (j = 1, \cdots, m)$$

$$d_j = -\frac{1}{n} \qquad (j = m + 1, \cdots, N).$$

Then the sequences D_N satisfy the condition W. If also the sequences A_{iN} satisfy the condition W, the distribution of $\bar{x}_i - \bar{x}'_i$ approaches the normal distribution as N increases, by the theorem of Section 2. If the joint distribution of $\bar{x}_1 - \bar{x}'_i$ and $\bar{x}_2 - \bar{x}'_2$ approaches a normal distribution with non-singular correlation matrix, the distribution of T'^2 approaches that of χ^2 with two degrees of freedom.

The correlation matrix of $(\bar{x}_1 - \bar{x}'_1)$ and $(\bar{x}_2 - \bar{x}'_2)$ will be of rank two in the limit if the correlation coefficient between $(\bar{x}_1 - \bar{x}'_1)$ and $(\bar{x}_2 - \bar{x}'_2)$ approaches a limit ρ, where $|\rho| < 1$. By (7.14) and (7.15) this is equivalent to saying that the absolute value of the angle between the vectors A'_{1N} and A'_{2N} is eventually greater than a positive lower bound. We shall show that, if the correlation coefficient approaches, as $N \to \infty$, a limit ρ whose absolute value is less than one, and if A_{1N} and A_{2N} satisfy the condition W, then $(\bar{x}_1 - \bar{x}'_1)$ and $(\bar{x}_2 - \bar{x}'_2)$ are *jointly* normally distributed in the limit.

Let δ_1 and δ_2 be any two real numbers not both zero. Then the sequence

$$A_N^* = (a_1^*, \cdots, a_N^*)$$

where

$$a_j^* = \delta_1 a_{1j} + \delta_2 a_{2j}$$

will be shown to satisfy the condition W. If either δ_1 or δ_2 is zero this is trivial; assume therefore that neither is zero. Without loss of generality we may assume that $\sum_{j=1}^{N} a_{ij} = 0$, for if this were not so we could replace the original a_{ij} by $a'_{ij} = a_{ij} - N^{-1} \sum_j a_{ij}$ as was done in Section 2. Let ρ' be such that $1 > \rho' > |\rho|$.

For N sufficiently large we have

$$\begin{aligned}\mu_2(A_N^*) &\geq N^{-1}(\delta_1^2 \sum_j a_{1j}^2 - 2 \mid \delta_1 \delta_2 \sum a_{1j} a_{2j} \mid + \delta_2^2 \sum_j a_{2j}^2) \\ &\geq N^{-1}(\delta_1^2 \sum_j a_{1j}^2 - 2\rho' \mid \delta_1 \delta_2 \mid \sqrt{(\sum_j a_{1j}^2)(\sum_j a_{2j}^2)} + \delta_2^2 \sum_j a_{2j}^2) \\ &= N^{-1}[(\mid \delta_1 \mid \sqrt{\sum_j a_{1j}^2} - \mid \delta_2 \mid \sqrt{\sum_j a_{2j}^2})^2 \\ &\qquad + 2(1 - \rho') \mid \delta_1 \delta_2 \mid \sqrt{(\sum_j a_{1j}^2)(\sum_j a_{2j}^2)}]\end{aligned}$$

and

$$\mu_2(A_N^*) \leq 2(\delta_1^2 \mu_2(A_{1N}) + \delta_2^2 \mu_2(A_{2N})).$$

Hence

$$\mu_2(A_N^*) = \Omega[\max \{\mu_2(A_{1N}), \mu_2(A_{2N})\}]. \tag{7.16}$$

Also $\mu_r(A_N^*)$ is a sum of constant multiples of terms of the type

$$N^{-1} \sum_j a_{1j}^i a_{2j}^{r-i}.$$

By Schwarz' inequality

$$N^{-1} \sum_j a_{1j}^i a_{2j}^{r-i} \leq N^{-1}(\sum_j a_{1j}^{2i})^{\frac{1}{2}}(\sum_j a_{2j}^{2(r-i)})^{\frac{1}{2}} = (\mu_{2i}(A_{1N})\mu_{2(r-i)}(A_{2N}))^{\frac{1}{2}}. \tag{7.17}$$

The required result follows from (7.16) and (7.17).

Since the sequences A_N^* satisfy the condition W, the limiting distribution of

$$\delta_1(\bar{x}_1 - \bar{x}_1') + \delta_2(\bar{x}_2 - \bar{x}_2'),$$

for any pair δ_1, δ_2 not both zero, is normal. From this and a theorem of Cramér and Wold ([11] Theorem 1; see also [8], Theorem 31) it follows that if the joint distribution of $(\bar{x}_1 - \bar{x}_1')$ and $(\bar{x}_2 - \bar{x}_2')$ approaches a limit, this limit must be the normal distribution. From a theorem of Radon ([12]; see also Cramér [8], page 101) it follows that if the joint distribution of $(\bar{x}_1 - \bar{x}_1')$ and $(\bar{x}_2 - \bar{x}_2')$ does not approach a limit as $N \to \infty$ it is possible to find two subsequences of the sequence $(1, 2, \cdots, N, \cdots$ ad inf.) for each of which the joint distribution approaches a different limit. This contradicts the previous result. Hence the limit exists and is the normal distribution. This proves our statement that the limiting distribution of T'^2 is the χ^2 distribution with two degrees of freedom.

The statistic T'^2 seems to be appropriate for testing the null hypothesis that two bivariate distributions Π_1 and Π_2 are identical if the alternatives are restricted to the case where Π_2 differs from Π_1 only in the mean values, i.e., the distribution Π_2 can be obtained from Π_1 by a translation. This is no restriction as compared with Hotelling's T-test since also the T-test is based on the assumption that the two normal populations differ at most in their mean values, i.e., the covariance matrices in the two populations are assumed to be equal.

REFERENCES

[1] A. Wald and J. Wolfowitz, *Annals of Math. Stat.*, Vol. 14 (1943), p. 378.
[2] Henry Scheffé, *Annals of Math. Stat.*, Vol. 14 (1943), p. 305.
[3] H. Hotelling and M. R. Pabst, *Annals of Math. Stat.*, Vol. 7 (1936), p. 29.
[4] E. J. G. Pitman, *Supp. Jour. Roy. Stat. Soc.*, Vol. 4 (1937), p. 225.
[5] E. J. G. Pitman, *Supp. Jour. Roy. Stat. Soc.*, Vol. 4 (1937), p. 119.
[6] E. J. G. Pitman, *Biometrika*, Vol. 29 (1938), p. 322.
[7] B. L. Welch, *Biometrika*, Vol. 29 (1937), p. 21.
[8] Harald Cramér, *Random Variables and Probability Distributions*, Cambridge, 1937.
[9] J. V. Uspensky, *Introduction to Mathematical Probability*, New York, 1937.
[10] H. Hotelling, *Annals Math. Stat.*, Vol. 2 (1931), p. 359.
[11] H. Cramér and H. Wold, *Journal London Math. Soc.*, Vol. 11 (1936), pp. 290–4.
[12] J. Radon, *Sitzungsberichte Akademie Wien*, Vol. 122 (1913), pp. 1295–1438.
[13] R. A. Fisher, *The Design of Experiments*, Edinburgh, 1937. (Especially Section 21.)
[14] R. A. Fisher, *Statistical Methods for Research Workers*, Edinburgh, 1936. (Especially Section 21.02.)