Journal of the Indian Statistical Association
Vol. 27. 1989, 1—8

SOME DISTRIBUTION-FREE TESTS FOR TESTING HOMOGENEITY OF LOCATION PARAMETERS AGAINST ORDERED ALTERNATIVES

Amita and **Subhash C. Kochar**
Department of Statistics,
Panjab University, Chandigarh, India

Abstract

Let X_{ij}, $j = 1, 2, \cdots n_i$ be independent random samples from populations with distribution functions $F_i(x) = F(x - \theta_i)$, $i = 1, 2, \cdots k$. The problem considered is to test the null hypothesis $H_0 : \theta_1 = \theta_2 = \cdots = \theta_k$ against the ordered alternative $H_1 : \theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_k$. The proposed test statistics are linear combinations of the consecutive two-sample statistics proposed by Deshpande and Kochar (1982) for the location problem. The optimum weights are obtained and asymptotic relative efficiencies have been computed.

1. Introduction

Let X_{ij} $(j = 1, 2, \cdots n_i)$ denote a random sample from the a absolutely continuous distribution $F_i(x)$ $(i = 1, 2, \cdots, k)$, the samples being independent. Let $N = \sum^k n_i$. We assume that F_i's differ only in their location alternatives, that is, $F_i(x) = F(x - \theta_i)$, $i = 1, 2, \cdots, k$. We wish to test the null hypotheis.

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_k \tag{1·1}$$

against the ordered alternative

$$H_1 : \theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_k \tag{1·2}$$

with at least one strict inequality.

This problem has been earlier studied by a number of researchers including Jonckheere (1954). Tryon and Hettmansperger (1973), Govindarajulu and Haller (1977), Shanubhogue and Gore (1985), among others.

Received : June 1987

We propose here a class of distribution free tests based on weighted linear combinations of consecutive two sample U-statistics for the above testing problem. The tests are introduced in the second section and their distributions are discussed in section 3. Assuming location parameters to be equally spaced and the sample sizes to be all equal, the optimal members in this class of tests are identified by obtaining the weighting coefficients which maximize the efficacy. In the last section, the proposed tests are compared with other known tests in the Pitman asymptotic relative efficiency sense.

2. The Proposed Tests

Under the alternative H_1, for $i < j$, the observations from the ith population tend to be smaller than those from the jth population. Deshpande and Kochar (1982) proposed the following class of distribution free tests for the two sample location problem.

Let c and d be two fixed integers such that $2 \leqslant c, d \leqslant \min(n_1, n_2, \cdots, n_k)$. Define for $i < j$, $i, j = 1, 2, \cdots, k$,

$$\Phi_{i,j}^{(c,d)} (X_{i\alpha_1}, \cdots, X_{i\alpha_c} ; X_{j\beta_1}, \cdots, X_{j\beta_d})$$

$$= \begin{cases} 2 & \text{if } \min(X_{i\alpha_1}, \cdots, X_{i\alpha_c}) \leqslant \min(X_{j\beta_1}, \cdots, X_{j\beta_d}) \\ & \text{and } \max(X_{i\alpha_1}, \cdots, X_{i\alpha_c}) \leqslant \max(X_{j\beta_1}, \cdots, X_{j\beta_d}) \\ 1 & \text{if either } \min(X_{i\alpha_1}, \cdots, X_{i\alpha_c}) \leqslant \min(X_{j1}, \cdots, X_{j\beta_d}) \\ & \text{or } \max(X_{i\alpha_1}, \cdots, X_{i\alpha_c}) \leqslant \max(X_{j1}, \cdots, X_{j\beta_d}) \\ 0 & \text{otherwise} \end{cases} \tag{2·1}$$

Let $U_{i,j}^{(c,d)}$ be the two-sample U-statistic associated with the above kernel and given by

$$U_{i,j}^{(c,d)} = \left[\binom{n_i}{c}\binom{n_j}{d}\right]^{-1} \sum_c \Phi_{i,j}^{(c,d)} (X_{i\alpha_1}, \cdots, X_{i\alpha_c} : X_{j\beta_1}, \cdots, X_{j\beta_d}) \tag{2·2}$$

where $\sum_c$ denotes summation extended over all combinations of c integers $(\alpha_1, \cdots, \alpha_c)$ chosen from $(1, \cdots, n_i)$ and all combinations of d integers

$(\beta_1, \cdots, \beta_d)$ chosen from $(1, \cdots, n_j)$. Clearly $U_{i,j}^{(1,1)}$ is the Mann-Whitney statistic computed for the ith and jth samples.

For testing H_0 against H_1, we propose the class of statistics

$$L_{c,d} = \sum_{i=1}^{k-1} a_i \; U_{i,\,i+1}^{(c,\,d)} \tag{2·3}$$

where $a_1, \cdots, a_{k-1}$ are some real constants to be chosen suitably. For the sake of convenience we shall omit the superscripts (c, d) from $\Phi_{i,j}^{(c,d)}$ and $U_{i,j}^{(c,d)}$. Obviously large values of $L_{c,d}$ are significant for testing H_0 against H_1 and the tests are distribution free.

3. Distribution of $L_{c,d}$

Let

$$\begin{aligned} \mu_{i,\,i+1} &= E\,[U_{i,\,i+1}] \\ &= d \int_{-\infty}^{\infty} F_i^c(x)\; F_{i+1}^{d-1}(x)\; dF_{i+1}(x) \\ &\quad + d \int_{-\infty}^{\infty} [\,1-\bar{F}_i^c(x)\,]\; \bar{F}_{i+1}^{d-1}(x)\; dF_{i+1}(x) \end{aligned} \tag{3·1}$$

where $\bar{F} = 1-F$.

Therefore,

$$E\,[\,L_{c,d}\,] = \sum_{i=1}^{k-1} a_i \; \mu_{i,\,i+1} \tag{3·2}$$

Under H_0,

$$E\,[\,L_{c,d}\,] = \sum_{i=1}^{k-1} a_i$$

Let

$$\mathbf{U}' = (U_{1,2}\; ;\; U_{2,3}\; ;\; \cdots\; ;\; U_{k-1,k}) \tag{3·3}$$

Since the $U_{i,j}$'s are two-sample U-statistics, the joint limiting normality of $\mathbf{U}$ follows immediately (see Lehmann (1963)) as stated below.

Theorem 3·1. *The asymptotic distribution of* $\sqrt{N}\,[\mathbf{U} - E(\mathbf{U})]$ *as* $N \to \infty$ *in such a way that* $n_i / N \to p_i$, $0 < p_i < 1$, *for* $i = 1, \cdots, k$ *is multivariate normal with mean vector* $\mathbf{O}$ *and disperasion matrix* $\Sigma = ((\sigma_{ij}))$,

where $N = \sum_{i=1}^{k} n_i$ *and*

$$\sigma_{i,j} = \begin{cases} \frac{c^2}{p_i}\, \xi^{(i)}_{i,i+1;\,i,i+1} + \frac{d^2}{p_{i+1}}\, \xi^{(i+1}_{i,i+1,i,i+1} & \textit{for } i=j=1,2,\cdots,k-1 \\ \frac{cd}{p_{i+1}}\, \xi^{(i+1)}_{i,i+1;\,i+1,i+2} & \textit{for } j=i+1, \quad i=1,2,\cdots,k-2 \\ \frac{cd}{p_i}\, \xi^{(i)}_{i-1,i;\,i,i+1} & \textit{for } j=i-1- \quad i=2,3,\cdots,k-1 \\ 0, & \textit{otherwise}, \end{cases} \tag{3·3}$$

$$\xi^{(i)}_{i,i+1;i,i+1} = E[\{\psi^{(i)}_{i,i+1}(X)\}^2] - E^2[U_{i,i+1}]. \tag{3·5}$$

$$\xi^{(i)}_{i,i+1;i,i+1} = E[\{\psi^{(i+1)}_{i,i+1}(X)\}^2] - E^2[U_{i,i+1}] \tag{3·6}$$

$$\xi^{(i)}_{i,i+1;i+1,i+2} = E[\psi^{(i+1)}_{i,i+1}(X)\,\psi^{(i+1)}_{i+1,i+2}(X)] \tag{3·7}$$

$$- E[U_{i,i+1}]\, E[U_{i+1,i+2}].$$

$$\psi^{(i)}_{i,j}(x) = E[\Phi_{i,j}(x, X_{i2}, \cdots X_{ic};\ X_{jl} \cdots, X_{jd})]. \tag{3·8}$$

$$\psi^{(j)}_{i,j}(x) = E[\Phi_{i,j}(X_{i1}, \cdots, X_{ic};\ x, X_{j2}, \cdots, X_{jd})]. \tag{3·9}$$

It can be seen that under H_0,

$$\sigma_{ij} = \begin{cases} 2c^2d^2\rho_m \left\{\frac{1}{p_i} + \frac{1}{p_i+1}\right\}, & \text{for } i=j=1,2,\cdots,k-1, \\ \frac{-2c^2d^2\rho_m}{p_i+1} & \text{for } j=i+1, i=1,2,\cdots,k-2, \\ \frac{-2c^2d^2}{p_i}\rho_m & \text{for } j=i-1, i=2,3,\cdots,k-1, \\ 0 & \text{Otherwise}; \end{cases} \tag{3·10}$$

where

$$m = c+d$$

and

$$\rho_m = \frac{1}{(m-1)^2 (2m-1)} \left[1 - \binom{2m-2-1}{m-1} \right] \tag{3·11}$$

In case all the sample sizes are equal, that is, $p_1 = p_2 \cdots = p_k = 1/k$, (3·10) becomes

$$\sigma_{i,j} = \begin{cases} 4\, c^2\, d^2 k\, \rho_m & \text{for } i=j=1, 2, \cdots, k-1 \\ -2\, c^2\, d^2 k\, \rho_m & \text{for } j=i+1,\ i=1,2,\cdots, k-2 \\ -2\, c^2\, d^2 k\, \rho_m & \text{for } j=i-1,\ i=2, 3,\cdots, k-1 \\ 0\ , & \text{otherwise} \end{cases}$$

Since $L_{c,\,d}$ is a linear combination of the components of the vector **U**, the proof of the following theorem follows immediately.

Theorem 3·2 : *The asymptotic distribution of* $\sqrt{N}\ [L_{c,\,d} - E\,(L_{c,d})]$ *as* $N \to \infty$ *in such a way that* $n_i / N \to p_i$, $0 < p_i < 1$, $i = 1, 2, \cdots, k$, *is normal with mean zero andvariance* $\mathbf{a}' \Sigma\, \mathbf{a}$;

Under H_0 $E\ [L_{c,\,d}] = \sum_{i=1}^{k-1} a_i$ *and* $\mathbf{a}' \Sigma\, \mathbf{a} = 4c^2 d^2 k\ _m[\sum_{i=1}^{k-1} a_i^2 - \sum_{i=1}^{k-2} a_i\, a_{i+1}]$ *in case the sample sizes are all equal.*

4. Optimal Choice of Weights

In this section, we consider the problem of obtaining the optimal weights, a_i's so that for fixed c and d, the test $L_{c,\,d}$ has maximum efficacy for the sequence of Pitman type alternatives

H_N : $F_i(x) = F\,[\,X - N^{-1/2}\,\theta_i\,]$, $i = 1, 2, \cdots, k$; F being an absolutely continuous distribution function with probability density function f.

For efficiency comparisons, we consider the equal sample size case and the equally spaced alternatives of the type, $\theta_i = i\,\theta$, $\theta > 0$ for $i = 1, 2, \cdots, k$. Then the alternative H_N becomes

$$H'_N : F_i(x) = F(x - N^{-1/2}\, i\theta) \tag{4·1}$$

The following theorem gives the asymptotic distribution of **U** under the sequence of alternatives H'_N. The proof is routine and hence omitted.

Theorem 4·1 : *Let X_{ij} be independent random variables with cumulative distribution function $F_i(x)$, $j = 1, 2, \cdots, n_i$, $i = 1, 2, \cdots, k$, as given by (4·1). The limiting distribution of $N^{1/2}\,[\mathbf{U} - \mathbf{E}_{k-1}]$ is $(k-1)$ dimensional multivariate normal with mean vector $\mu\,\mathbf{E}_{k-1}$ and dispersion matrix $\Sigma = ((\sigma_{ij}))$ given by (3·12); where*

$$\mu = c\,d\,\theta \int_{-\infty}^{\infty} \left[F^{m-2}(x) + \bar{F}^{m-2}(x) \right] f(x)\, d\dot{F}(x) \tag{4·2}$$

$$m = c + d$$

and

$$\mathbf{E}_{k-1} = [\,1\,]_{(k-1)\times 1}$$

Proceeding as in Kochar and Gupta (1986), one can easily prove the following theorem.

Theorem 4·2 : *Under the sequence of alternative $\left\{ H'_N \right\}$ given by (4·1), the efficacy of the test $L_{c,d}$ is maximized when*

$$a^*_i = -\frac{i\,(k-1)}{2k}, \quad i = 1, 2, \cdots, k-1,$$

and the efficacy of the $L^*_{c,d}$ test with these optimal choice of weights is

$$e\left(L^*_{c,d}\right) = \frac{(k-1)^2\, H^2_m}{24\,\rho_m},$$

where

$$H_m = \int_{-\infty}^{\infty} \left[F^{m-2}(x) + F^{m-2}(x) \right] f(x)\, d\,F(x)$$

Clearly, $e(L^*_{c,d})$ depends on c and d only through the sum $m = c + d$.

5. Asymptotic Relative Effciencies

Here, we compare the asymptotic relative efficiencies of the $L^*_{c,d}$ tests relative to some other known tests for this testing problem.

The asymptotic relative efficiency of the $L^*_{c,d}$ test with respect to the Jonkheer's test (1954) is

$$e\left(L^*_{c,d}, J\right) = \frac{H_m^2}{24\,\rho_m}\left[\int_{-\infty}^{\infty} f^2(x)\,dx\right]^{-2} \tag{5·1}$$

which is the same as the ARE of the corresponding two-sample test with respect to the Wilcoxon-Mann-Whitney test as given by Deshpande' and Kochar (1982).

Table 5·1 gives these values for some standard distributions.

Table 5·1

ARE's of the $L^*_{c,d}$ test relative to Jonckheere test

Distn. m	4	5	6	7	8	9	10
Double Exp.	·9402	·85597	·7677	·6863	·6155	·5554	.5047
Logistic	·9947	.9739	.9391	·8964	·8509	·80599	·7632
Normal	1·0143	1·0203	1·0273	1·0049	·9814	·9674	·9355
Exponential	1·2281	1·5217	1·8406	2·1690	2·5007	2·8336	3·1667

For $m=2$ and $m=3$, the ARE's are 1 for all distributions.

Govindarajulu and Haller (1977) also proposed a class of distribution-free tests for this problem. This class includes equivalent of tests proposed by Bhapkar (1961) and Deshpande' (1965). Pitman ARE of the $L^*_{c,d}$ test with corresponding to Deshpande' test is

$$e\left(L^*_{c,d}, L\right) = \frac{\rho_k\, H_m^2}{\rho_m\, H_k^2}$$

Clearly, for $m = k$, $e\,(L^*_{c,d}, L) = 1$, Table 5·2 gives the values of $e\,(L^*_{c,d}, L)$ for some standard distribution.

Table 5·2

ARE's of $L^{*}_{c,\,d}$ tests relative to Deshpande's L–Test.

Distribution	m						
Double exponential	5	·9104	1·1144	1·3908	1·6960	2·0116	2·3307
Logistic	10	·7672	·8126	·8968	1	1·1122	1·2293
Normal	10	·9355	·9169	·9532	1	1·0618	1·1319

It is clear from the above tables that the newly proposed tests are quite efficient.

References

1. Bhapkar, V. P. (1961). A nonparametric test for the problem of several samples, *Ann. Math. Statist.* **3,** 1108–1117.

2. Deshpande, J. V. (1965). A nonparametric test based on U–statistics for the problem of several samples *J. Indian Statist. Assoc.* **3,** 20–29.

3. Deshpande, J. V. and Kochar, S. C. (1982). Some competitors of the Wilcoxon-Mann-Whitney test for the location alternative. *J. Indian Statist. Assoc.* **20,** 9–18.

4. Graybil, F. A. (1969). *Introduction to Matrices with Applications in Statistics,* Wordsworth Publishing Company, California.

5. Govindarajulu, Z. and Haller, H. S. (1977). c-sample tests of homogeneity against ordered alternative. Proceedings of the symposium to honour Jerzy Neymann.

6. Jonckheere, A. R. (1954): A distribution-free k samples test against ordered alternatives. *Biometrika* **41,** 133–145.

7. Kochar, S. C. and Gupta, R. P. (1985). A class of distribution-free tests for testing homogeneity of variances against ordered alternatives. Proceedings of the Symposium on Order Restricted Statistical Inference held in Iowa City.

8. Puri, M. L. (1985). Some distribution-free k sample rank order tests of homogeneity against ordered alternatives, *Comm. Pure, Appl. Math.* **18,** 51–53.

9. Tryon, P. V. and Hettmansperger, T. P. (1973). A class of nonparametric tests for homogeneity against ordered alternative. *Ann. Statist.* **1,** 1061–1070.