Ann. Inst. Statist. Math.
34 (1982), Part A, 327-334

NONPARAMETRIC TESTS FOR HOMOGENEITY OF SCALE AGAINST ORDERED ALTERNATIVES

K. S. MADHAVA RAO

(Received Oct. 23, 1981)

Summary

In this paper the nonparametric several sample scale problem is considered and some tests are proposed for the hypothesis of homogeneity versus ordered alternatives. These tests are based on statistics that are weighted linear combinations of Sugiura (1965, *Osaka J. Math.*, 2, 385–426) type statistics proposed for testing homogeneity of scale against the omnibus alternative. For each class of test statistics suggested, the member with maximum Pitman efficiency is identified. The optimal statistics are compared with their parametric and nonparametric competitors.

1. Introduction

Let X_{ij}, $j=1,\cdots,n_i$ be independent random variables distributed according to the common cumulative distribution function $F_i(x)$, where

$$F_i(x)=F((x-\theta_i)/\sigma_i)\ , \tag{1.1}$$

and θ_i is the location parameter and $\sigma_i>0$ is the scale parameter ($i=1,\cdots,c$). The case $\sigma_i=1$ for every $i=1,\cdots,c$ corresponds to the several sample location problem. Literature on nonparametric tests of homogeneity against the ordered location alternatives for several sample problem is quite extensive (e.g. [3], [4], [6], [7], [8]). Some of these tests have been discussed in [2]. However, the problem of testing the hypothesis of homogeneity of populations (where we assume that $\theta_i=\theta$, $i=1,\cdots,c$) against the ordered scale alternatives does not seem to have received much attention. These hypotheses, which are of interest in this paper, could be formally stated as follows:

AMS 1979 subject classifications: Primary 62G10; Secondary 62J10.
Key words and phrases: Ordered scale alternatives, Pitman efficiency, Weighted rank statistics.

(1.2) $$H_0: \sigma_1=\sigma_2=\cdots=\sigma_c$$

against the alternatives

(1.3) $$H_1: \sigma_1\leqq\sigma_2\leqq\cdots\leqq\sigma_c ,$$

where at least one inequality is strict. We assume that $\theta_i=\theta$, $i=1, \cdots, c$ in (1.1) and when using ranks, without loss of generality we can assume $\theta=0$. Recently, Govindarajulu and Haller [6] proposed a class of test statistics for testing H_0 versus H_1, which are weighted sums of rank statistics, the weights being optimal in the sense of Pitman efficiency. Some other tests for this problem such as a locally most powerful rank test, a parametric test based on the "likelihood derivave" and a heuristic class of rank tests are due to Govindarajulu and Gupta [5].

Our particular interest in this note is the work of Sugiura [10], that proposes two statistics for testing the hypothesis of homogeneity of scale against omnibus alternative. It is shown in this article that a modification of Sugiura type statistics will yield a class of weighted linear combination of statistics, that are sensitive to H_1. The optimal member (in the Pitman asymptotic efficiency sense) in the class of statistics considered are derived. It turns out that these statistics have asymptotic efficiencies similar to those considered elsewhere.

2. Proposed class of test statistics

Define for $i=1,\cdots,c$

(2.1) $$\phi_1^{(i)}(X_1,\cdots,X_c) = \begin{cases} \dfrac{(j-1)_r}{(c-1)_r}+\dfrac{(c-j)_s}{(c-1)_s}, & \text{if } X_i \text{ has rank } j \text{ among } cX\text{'s} ; \\ 0 , & \text{otherwise} , \end{cases}$$

where $0\leqq r, s\leqq c-1$ except for $(r, s)=(0, 0)$ and $(1, 1)$, $(K)_r=K(K-1)\cdots(K-r+1)$, $(K)_0=1$, and

(2.2) $$\phi_2^{(i)}(X_{11}, X_{12};\cdots; X_{c1}, X_{c2}) = \begin{cases} 1 , & \text{if } X_{i1}<X_{km}<X_{i2} \text{ or } X_{i2}<X_{km}<X_{i1} \\ & \qquad \text{for all } k\neq i \text{ and } m=1, 2 ; \\ 0 , & \text{otherwise} . \end{cases}$$

Let

(2.3) $$U^{(i)}=(n_1\cdots n_c)^{-1} \sum_{\alpha_1=1}^{n_1}\cdots\sum_{\alpha_c=1}^{n_c} \phi_1^{(i)}(X_{1\alpha_1},\cdots,X_{c\alpha_c}) ,$$

and

$$(2.4) \qquad V^{(i)}=\left[\binom{n_1}{2}\cdots\binom{n_c}{2}\right]^{-1} \sum_{\alpha_1<\beta_1}\cdots\sum_{\alpha_c<\beta_c} \phi_2^{(i)}(X_{1\alpha_1}, X_{1\beta_1};\cdots; X_{c\alpha_c}, X_{c\beta_c}) ,$$

where $\sum_{\alpha_1<\beta_1}\cdots\sum_{\alpha_c<\beta_c}$ means the summation extending on all possible pairs (α_i, β_i) such that $1\leqq\alpha_i<\beta_i\leqq n_i$ for $i=1,\cdots, c$.

Now if H_0 is true, the events that the random variable X_i obtains rank j among $X_1,\cdots, X_c$ for $j=1,\cdots, c$ are equally likely. Therefore it readily follows that

$$(2.5) \qquad \mathrm{E}\,[\phi_1^{(i)}(X_1,\cdots, X_c)]=\frac{1}{c}\sum_{j=1}^{c}\left[\frac{(j-1)_r}{(c-1)_r}+\frac{(j-1)_s}{(c-1)_s}\right] ,$$

so that

$$(2.6) \qquad \mathrm{E}\,(U^{(i)})=\frac{1}{r+1}+\frac{1}{s+1} ,$$

since $\sum_{j=1}^{c}(j-1)_r=(c)_{r+1}/r+1$, and

$$(2.7) \qquad \mathrm{E}\,(V^{(i)})=\frac{1}{c(2c-1)} .$$

Motivated by the fact that if H_1 were true, we expect $U^{(i)}$ and $V^{(i)}$ to increase with i, we propose two class of statistics defined by

$$(2.8) \qquad K_1=\sum_{i=1}^{c} a_i U^{(i)} ,$$

$$(2.9) \qquad K_2=\sum_{i=1}^{c} b_i V^{(i)} .$$

It is assumed that the weights $a_i\,(b_i)$ are not all equal and are real constants. A particular member of the class of statistics $K_1\,(K_2)$ is identified by specifying $a_i\,(b_i)$, r and s. With each member of the class we can associate a test of H_0 which rejects H_1 at a significance level α if $K_1\,(K_2)$ exceeds some predetermined constant $K_1(\alpha)$ $(K_2(\alpha))$.

3. Asymptotic distributions

We first notice that $U^{(i)}\,(V^{(i)})$ is a U-statistic corresponding to the kernel $\phi_1^{(i)}\,(\phi_2^{(i)})$ generalized to the case of c samples. Let $N=\sum_{i=1}^{c} n_i$ tend to infinity in such a way that $n_i/N\to p_i$, where $0<p_i<1$, $i=1,\cdots, c$ are constants. Then following [10], the c-dimensional statistics $(U^{(1)},\cdots, U^{(c)})$ $((V^{(1)},\cdots, V^{(c)}))$, when properly standardized is distributed asymptotically according to c-variate normal distribution with mean vector zero and a certain covariance matrix. Thus the asymptotic normality of $K_1\,(K_2)$ is immediate, by noting that $K_1\,(K_2)$ is a linear combination of $U^{(i)}\,(V^{(i)})$. It is easy to verify that under H_0,

$$\text{(3.1)}\qquad \mathrm{E}_0\,(K_1)=\sum_{i=1}^{c} a_i\left(\frac{1}{r+1}+\frac{1}{s+1}\right),$$

$$\text{(3.2)}\qquad \mathrm{V}_0\,(K_1)=\left[c^2\frac{A(r,s)}{(c-1)^2}\right]\sum_{i=1}^{c}(a_i-\bar{a})^2/p_i\,,$$

and

$$\text{(3.3)}\qquad \mathrm{E}_0\,(K_2)=\sum_{i=1}^{c} b_i\left(\frac{1}{c(2c-1)}\right),$$

$$\text{(3.4)}\qquad \mathrm{V}_0\,(K_2)=\left[\frac{4c^2A(2c-1,\,2c-1)}{(c-1)^2(2c-1)^2}\right]\sum_{i=1}^{c}(b_i-\bar{b})^2/p_i\,,$$

where $c\bar{a}=\sum a_i$, $c\bar{b}=\sum b_i$,

$$\text{(3.5)}\qquad A(r,s)=\frac{r^2}{(2r+1)(r+1)^2}+\frac{s^2}{(2s+1)(s+1)^2}-\frac{2}{(r+1)(s+1)}+2B(r+1,\,s+1)\,,$$

$B(p,q)$ is the usual Beta function.

4. Optimal choice of weights

In order to assess the merits of the class of test statistics proposed here, the asymptotic efficiency will be computed relative to other known parametric and nonparametric competitors. Towards this, we first discuss the efficacy of the proposed class of statistics and derive the optimal weighting coefficients $a_i\,(b_i)$ such that the corresponding Pitman efficacy is maximized. We shall consider the following sequence of " near " alternatives:

$$\text{(4.1)}\qquad H_N:\ F_i(x)=F(x/(\sigma+N^{-1/2}d_i))$$

$i=1,\cdots,c$, where σ and d_i are some real constants (not all d_i's are equal), F is an absolutely continuous distribution function. Since (2.8) and (2.9) remain invariant if all the variables X_{ij}, $i=1,\cdots,c$, $j=1,\cdots,n_i$ are subjected to a change of scale, without loss of generality we may assume $\sigma=1$. Further, we consider the equal sample case i.e., $p_i=1/c$ is fixed, and the equally spaced scale alternatives of the type $d_i=id$, $d>0$ for $i=1,\cdots,c$. We now require the following two lemmas, which are relevant particular cases of results given in [10].

LEMMA 4.1. *Assume the sequence of distributions given by* (4.1) *of independent random variables* X_{ij}, $j=1,\cdots,n_i$ *for each index* $n=1,2,\cdots$, *where* $n_i=Np_i$ *with* p_i *a positive number* $(0<p_i<1)$, $i=1,2,\cdots,c$. *Suppose further that* F *possess a continuous derivative* f *and there exists a function* g *such that*

$$\left|\frac{1}{n}[f(y+h)-f(y)]\right|\leqq g(y) \qquad \text{and} \qquad \int_{-\infty}^{\infty} g(y)f(y)dy<\infty .$$

If $N=\sum_{i=1}^{c} n_i$ *and* $U'=[U^{(1)},\cdots,U^{(c)}]$, $L=[1]_{1\times c}$, *then as* $N\to\infty$, *the limiting distribution of* $N^{1/2}(U-[1/(r+1)+1/(s+1)]L)$ *is a c-variate normal with mean vector* $\mu'=(\mu_1,\cdots,\mu_c)$ *and the variance-covariance matrix* $\Sigma=(\sigma_{ij})_{c\times c}$, $i,j=1,\cdots,c$, *where*

$$\mu_i=(c-1)^{-1}c\left(i-\frac{c+1}{2}\right)dI_1 , \tag{4.2}$$

$$I_1=\int_{-\infty}^{\infty} xf(x)\{r[F(x)]^{r-1}-s[1-F(x)]^{s-1}\}dF(x) ,$$

and

$$\sigma_{ij}=\frac{A(r,s)}{(c-1)^2}c^3\left(\delta_{ij}-\frac{1}{c}\right) . \tag{4.3}$$

δ_{ij} *is the usual Kronecker delta and* $A(r,s)$ *is given by* (3.5).

LEMMA 4.2. *Let* $V'=[V^{(1)},\cdots,V^{(c)}]$. *Then under the assumptions of Lemma* 4.1, *the limiting distribution of* $N^{1/2}(V-[c(2c-1)]^{-1}L)$ *is a c-variate normal with mean vector* $\tilde{\mu}'=(\tilde{\mu}_1,\cdots,\tilde{\mu}_c)$ *and covariance matrix* $\tilde{\Sigma}=(\tilde{\sigma}_{ij})_{c\times c}$ *where*

$$\tilde{\mu}_i=2(c-1)^{-1}c\sum_{i=1}^{c}\left(i-\frac{c+1}{3}\right)dI_2 , \tag{4.4}$$

$$I_2=\int_{-\infty}^{\infty} xf(x)\{[F(x)]^{2c-2}-[1-F(x)]^{2c-2}\}dF(x) ,$$

and

$$\tilde{\sigma}_{ij}=\frac{4A(2c-1,2c-1)c^3}{(c-1)^2(2c-1)^2}(\delta_{ij}-1/c) . \tag{4.5}$$

Next we shall consider the problem of determining the optimal weights $a_i\,(b_i)$ associated with the class of statistics $K_1\,(K_2)$. The following theorem gives the optimal weights $a_i\,(b_i)$'s that maximize the efficacy of the statistics $K_1\,(K_2)$.

THEOREM 4.1. *Under the Assumptions of Lemma* 4.1 (4.2), *the efficacy of the test statistic* $K_1\,(K_2)$ *is maximized if*

$$a_i\,(b_i)=\left(i-\frac{c+1}{2}\right)\,,\qquad i=1,\cdots,c\,.$$

PROOF. We first recall that the efficacy of a test statistic T_N is defined as

$$e(T_N)=[\mu'(T_N)/\theta=0]^2/\sigma_0^2(T_N)\;, \tag{4.6}$$

where $\theta=dN^{-1/2}$, $d>0$. We have from (2.8) and (4.2),

$$\mathrm{E}_\theta\,(K_1)=\sum_{i=1}^{c} a_i\left\{c(c-1)^{-1}\left(i-\frac{c+1}{2}\right)\theta I_1+\left(\frac{1}{r+1}+\frac{1}{s+1}\right)\right\}\,. \tag{4.7}$$

The derivative of (4.7) with respect to θ, evaluated at $\theta=0$ is

$$c(c-1)^{-1}\sum_{i=1}^{c} a(i)\left(i-\frac{c+1}{2}\right)I_1\,. \tag{4.8}$$

From (3.2), (4.8) and (4.6) we obtain the efficacy of K_1, given by

$$e(K_1)=\frac{I_1^2G}{cA(r,s)}\,, \tag{4.9}$$

where

$$G=\frac{\left[\sum_{i=1}^{c} a_i(i-(c+1)/2)\right]^2}{\sum_{i=1}^{c}(a_i-\bar{a})^2}\,. \tag{4.10}$$

In (4.9) we need to maximize only G since other factors do not depend on a_i. Let us define $a'=(a_1,\cdots,a_c)$, $b'=(1-(c+1)/2,\cdots,c-(c+1)/2)$. Further let $M=(m_{ij})_{c\times c}$, $i,j=1,\cdots,c$ where $m_{ij}=\delta_{ij}-1/c$. Then G can be written as $(a'b)^2/a'Ma$. It is easily seen that M is singular and of rank $(c-1)$. Further observe that $\sum_i m_{ij}=\sum_j m_{ij}=\sum_i b_i=0$. We may assume without loss of generality that $a'J=0$ for $J=(1,1,\cdots,1)'$. Now it is easy to verify the following: (i) $a'b=a_0'b_*$ (ii) $a'Ma=a_0'M^*a_0$ where $a_0'=(a_1,\cdots,a_{c-1})$, $b_*'=(1-c,2-c,\cdots,c-1-c)$ and $M^*=\delta_{ij}+1$, $i,j=1,\cdots,c-1$. Clearly M^* is positive definite. Thus we have $G=(a'b)^2/a'Ma=(a_0'b_*)^2/a_0'M^*a_0$. Now using a well known result in Matrix algebra (e.g., see [9], p. 48); we immediately obtain the optimal weights

$$a_i=\left(i-\frac{c+1}{2}\right)\,,\qquad i=1,\cdots,c\,, \tag{4.11}$$

which completes the proof of the theorem. The proof for the class K_2 is exactly analogous. As an immediate consequence, we have the following corollary.

COROLLARY 1. *The optimum efficacy of* K_1 (K_2) *with* a_i $(b_i)=(i-(c+1)/2)$, *is given by*

$$e(K_1)=\frac{(c^2-1)I_1^2}{12A(r, s)}, \tag{4.12}$$

$$e(K_2)=\frac{(c^2-1)(2c-1)^2 I_2^2}{12A(2c-1, 2c-1)}. \tag{4.13}$$

5. Efficiency comparisons and conclusion

In this section we shall compare the asymptotic efficiencies of tests K_1 and K_2 relative to the parametric and nonparametric competitors. If we let $r=s=c-1$ and $r=s=2$ in (2.1) then the corresponding statistics $\tilde{K}_1$ and K_1^* can be considered as modified versions of the Bhapkar and Deshpande [1] D statistic and the Sugiura [10] D_{22} statistics suitable for testing H_0 versus H_1. The asymptotic efficiencies of K_1 relative to K_1^* can be shown to be the same as exhibited in Tables 6 through 10 of Sugiura [10].

As stated earlier Govindarajulu and Gupta [5] developed a locally most powerful rank statistic S_{1N}, a statistic S_{2N} based on the "likelihood derivative" and a class of weighted sum of Chernoff-Savage type statistics S_{3N} for testing H_0 versus H_1. It can be seen that $e(K_1^*)=15(c^2-1)\left\{\int_{-\infty}^{\infty} xf^2(x)[2F(x)-1]dx\right\}^2$, and K_1^* is as efficient as the statistic S_{3N}, specialized to Mood type scores considered by Govindarajulu and Gupta [5]. The efficacies of the statistics K_1^*, S_{1N} and S_{2N} for Normal (0, 1) and Exponential (0, 1) error distributions are given by Table 2 of Govindarajulu and Gupta [5]. The efficiencies of the test K_2 relative

Table 1 Asymptotic efficiency of K_2 relative to S_{1N}

c	2	3	4	5	6	7	8	9	10
Normal	0.760	0.812	0.864	0.898	0.918	0.928	0.931	0.929	0.925
Double Exponential	0.899	0.935	0.961	0.970	0.967	0.957	0.943	0.927	0.911
Exponential	0.784	0.834	0.885	0.913	0.930	0.936	0.936	0.936	0.924

Table 2 Asymptotic efficiency of K_2 relative to S_{2N}

c	2	3	4	5	6	7	8	9	10
Normal	0.304	0.464	0.576	0.653	0.706	0.742	0.767	0.782	0.793
Double Exponential	1.996	2.964	3.557	3.914	4.127	4.250	4.309	4.337	4.332
Exponential	0.320	0.486	0.602	0.677	0.730	0.764	0.786	0.804	0.808

to S_{1N} and S_{2N} for Normal (0, 1), Exponential (0, 1) and Double Exponential (0, 1) error distributions are given in Table 1 and Table 2 respectively. Further it can be shown that the efficiency of K_1^* relative to S_{1N} for these three distributions is respectively 0.760, 0.784 and 0.899. From these values we can see that the test K_2 performs better than the test K_1^*. Further the K_2 test has reasonably good efficiency compared to S_{1N}, while it turns out to be considerably efficient as compared to S_{2N}, for double exponential residuals.

Acknowledgements

The author is deeply indebted to Dr. A. P. Gore for his suggestion to investigate this problem and also for his constant help. Thanks are also due to the University Grants Commission, New Delhi (India) for the award of Teacher Fellowship under the Faculty Improvement Programme.

UNIVERSITY OF POONA*

REFERENCES

[1] Bhapkar, V. P. and Deshpande, J. V. (1968). Some nonparametric tests for multisample problems, *Technometrics*, **10**, 578-585.
[2] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*, John Wiley and Sons, New York.
[3] Chacko, V. J. (1963). Testing homogeneity against ordered alternatives, *Ann. Math. Statist.*, **34**, 945-956.
[4] Deshpande, J. V. (1978/79). Tests based on *c*-plets of observations for homogeneity against ordered alternatives, *S. C. Das Memorial Volume*, published by Utkal University, India, 19-25.
[5] Govindarajulu, Z. and Gupta, G. D. (1978). Tests for homogeneity of scale against ordered alternatives, *Trans. of the 8th Prague Conference on Information Theory, Statistical Decision Functions, Random Processess* (ed. J. Koželnik, et al.), Academia Publishing House, Prague, Volume A, 235-245.
[6] Govindarajulu, Z. and Haller, H. S. (1977). *c*-sample tests of homogeneity against ordered alternatives, *Proceedings of the symposium to honour Jerzy Neyman* (ed. R. Bartoszynski, et al.), Polish Scientific Publishers, Warszawa, 91-102.
[7] Jonckheere, A. R. (1954). A distribution-free *k*-sample test against ordered alternatives, *Biometrika*, **41**, 133-145.
[8] Puri, M. L. (1965). Some distribution-free *k*-sample rank tests of homogeneity against ordered alternatives, *Commun. Pure Appl. Math.*, **18**, 51-63.
[9] Rao, C. R. (1966). *Linear Statistical Inference and its Applications*, Second Printing, John Wiley and Sons, New York.
[10] Sugiura, N. (1965). Multisample and multivariate nonparametric tests based on *U* statistics and their asymptotic efficiencies, *Osaka J. Math.*, **2**, 385-426.

* On study leave from M.G.M. College, Udupi, (D.K.) 576102, India.