⟨**Research Paper**⟩

Journal of the Korean
Statistical Society
Vol. 13, No. 1, 1984

An Adaptive Distribution-Free Test for the Multi-Sample Location Problem

Il Seong Song*

ABSTRACT

An adaptive distribution-free test is proposed for testing the equality of k independent distributions against unrestricted alternatives. In this paper, several rank-sum test statistics are considered as the components of the adaptive one. The emprical powers of the adaptive testing procedure are compared to those of the classical F test and the component tests through a Monte Carlo study. The results show that the adaptive test has good power properties over a wide class of underlying distributions.

1. Introduction

In statistical testing procedure, it is not uncommon to have few ideas of the underlying distribution before analyzing data. In this situation the power of the test may be improved by employing an adaptive rule. That is, the data may be used first to select the model which seems most appropriate and then to make a statistical inference based on the model chosen. In such an adaptive testing procedure, the overall significant α-level should be maintained. These facts suggest we construct adaptive testing procedures based on truly nonparametric distribution-free tests in such a way that the preliminary selection of the model is statistically independent of these final tests.

In this paper, we deal with an adaptive distribution-free test for multi-sample location problem. Let $X_{ij}(i=1, \ldots, k;\ j=1, \ldots, n_i)$ be k independent random samples from continuous distributions with cdf's $F(x-\Delta_1), \ldots,\ F(x-\Delta_k)$, respectively, where Δ_i denotes the location shift parameter of the ith population. We want to test, on the basis of the above samples, the hypothesis that $\Delta_1=\ldots=\Delta_k$ against unrestricted alternatives. To construct an

* Department of Statistics, Sungshin Women's University, Seoul 132, Korea

adaptive testing procedure for the settings we propose several distribution-free tests based on rank statistic whose performances depend on the tailweight and amount of skewness in the underlying distribution.

An adaptive distribution-free test for one- and two-sample location problems have been developed by Randles and Hogg(1973) and Hogg, Fisher and Randles (1975). Hogg, Fisher and Randles(1975) investigated a two-sample adaptive distribution-free test for the broad class of underlying distributions including symmetric and asymmetric ones.

For the multi-sample location problem Kruskal and Wallis(1952) proposed a test of nonparametric nature which is a direct generalization of the two-sample Wilcoxon rank test. Puri(1964) presented generalized rank tests for the multi-sample location problem to allow for arbitrary scores, as considered in detail for the two-sample setting in Randles and Wolfe (1979). Puri(1964) also derived the limiting distribution of the generalized version under the contiguous alternatives and their asymptotic relative efficiencies.

Some of the results contained in this paper have their origins in the paper of Hogg, Fisher and Randles(1975) wherein similar notions and results are given for the two-sample problem. A formulation of appropriate distribution-free rank tests is based on the generalized version of Puri(1964).

In Section 2, several rank tests with appropriate scores are proposed for the adaptive procedures and their asymptotic relative efficiencies are evaluated and summarized in Table 2.1. Section 3 contains the formulation of the adaptive selection scheme using skewness and kurtosis indicators. The result of a Monte Carlo study on the adaptive procedure is presented in Section 4. The results in this paper show that an adaptive procedure has improved the power over a wide class of distributions.

2. Test Statistics Employed for the Adaptive Rule

Before introducing the related test statistics, we state a few assumptions and notations which are mainly adaptations of those of Puri(1964). Let $X_{i1}, \ldots, X_{in_i}$ be the random sample from a population with continuous cdf $F^{(i)}(x)$, $i=1, \ldots, k$.

Let $N=\sum_{i=1}^{k} n_i$ and $\lambda_i = n_i/N$ and assume that for all N, the inequalities $0<\lambda_0 \leqslant \lambda_1, \ldots, \lambda_k \leqslant 1-\lambda_0 < 1$ hold for some fixed $\lambda_0 \leqslant 1/k$. Let $F_{n_i}{}^{(i)}(x) = (\text{number of } X_{ij} \leqslant x,\ j=1, \ldots, n_i)/n_i$ be the empirical distribution function of the n_i observations in the ith set. Define the combined empirical distribution function $H_N(x) = \lambda_1 F_{n_1}{}^{(1)}(x) + \cdots + \lambda_k F_{n_k}{}^{(k)}(x)$. The combined population cdf is $H(x)=\lambda_1 F^{(1)}(x)+\cdots+\lambda_k F^{(k)}(x)$. Let $I_N=\{x; 0<H_N(x)<1\}$. Then I_N is a random interval, given by $I_N=[X_{(1)}, X_{(N)})$, where $X_{(1)}\leqslant\cdots\leqslant X_{(N)}$ are order statistics of the N observations.

Consider again k independent random samples $X_{ij}(i=1, \ldots, k; j=1, \ldots, n_i)$ with continuous cdf's $F^{(i)}(x)=F(x-\Delta_i)$, respectively. We wish to test $H_0: \Delta_1=\ldots=\Delta_k$ against unrestricted alternatives.

For the above setting, Puri(1964) considered test statistics of the following form

$$R=\sum_{i=1}^{k}\{\sqrt{n}_i(T_{N,i}-\mu_{N,i})/A_N\}^2, \tag{2.1}$$

with critical region consisting of upper values of R, where n_i is the size of the ith sample,

$$\mu_{N,i}=E(T_{N,i}),$$

$$A_N^2=\int_0^1 \phi_N^2(u)\,du-\left(\int_0^1 \phi_N(u)du\right)^2, \tag{2.2}$$

and $T_{N,i}$ as the kernel of test statistic R is represented as follows.

$$T_{N,i}=(1/n_i)\sum_{j\in s_i}\phi_N(R_j^*/N), \tag{2.3}$$

where $\{s_1, \ldots, s_k\}$ is a partition of $\{1, \ldots, N\}$, $n_i=\text{card } s_i$, $N=\sum_{i=1}^{k} n_i$ and $R^*=(R_1^*, \ldots, R_N^*)$ is the rank vector of the variables $X_{ij}(i=1, \ldots, k;\ j=1, \ldots, n_i)$. In the above expressions, the $\phi_N(\cdot)$ functions must satisfy the following assumptions which are the conditions of Lemma 5.1 of Puri (1964) and we refer them as "score functions".

Assumption 2.1.

i) $\phi(H)=\lim_{N\to\infty}\phi_N(H)$ exists for $0<H<1$ and is not constant,

ii) $\int_{I_N}[\phi_N(H_N)-\phi(H_N)]dF_{n_j}^{(j)}(x)=o_p(1/\sqrt{N})$,

iii) $\phi_N(1)=o(\sqrt{N})$,

iv) $|\phi^{(i)}(H(x))|=|d^i\phi(H)/dH^i|\leqslant k(H(1-H))^{-i-\frac{1}{2}+\delta}$, for $i=0, 1, 2$, and for some $\delta>0$, and almost all x, where K is a constant which may depend on ϕ_N but will not depend on $F^{(1)}, \ldots, F^{(k)}$, $n_1, \ldots, n_k$ and N.

Thus $T_{N,i}$ is the average of the scores associated with the joint ranks of the ith sample. We are frequently at liberty to choose the scores so as to achieve desirable power properties which is discussed later in this section.

Under certain regularity conditions including Assumption 2.1(see Puri (1964)), the test statistic R has the limiting noncentral $\chi^2(k-1)$ under contiguous alternatives and the limiting $\chi^2(k-1)$ under H_0 and the asymptotic relative efficiency of an arbitrary R test with respect to the Kruskal-Wallis test is

$$ARE(R, KW)=(\int_{-\infty}^{\infty}\frac{d}{dx}\{\phi[F(x)]\}f(x)dx/\int_{-\infty}^{\infty}f^2(x)dx)^2/(12A^2), \quad (2.4)$$

where

$$A^2=\int_0^1\phi^2(u)du-(\int_0^1\phi(u)du)^2,$$

and f is the density of F.

We can demonstrate a wide variety of scores so as to improve the asymptotic relative efficiencies corresponding to specified underlying distributions. In this paper, to proceed the adaptive testing procedures, we consider the following five set of scores.

(1) Kruskal-Wallis scores(denoted by KW) are

$$\phi_N(i/N)=i/N, \quad i=1, \ldots, N \quad (2.5)$$

Hájek and Sidák(1967) show these scores are asymptotically efficient in detecting shifts in logistic type distribution.

(2) Light-tailed distribution scores(denoted by L) are

$$\phi_N(i/N)=\begin{cases}\frac{i}{N}-\frac{1}{4}, & i<\frac{N}{4}\\ 0, & \frac{N}{4}\leqslant i\leqslant\frac{3}{4}N\\ \frac{i}{N}-\frac{3}{4}, & i>\frac{3}{4}N\end{cases} \quad (2.6)$$

These scores are effective for detecting a shift in light-tailed distribution.

(3) Heavy-tailed distribution scores(denoted by H) are

$$\phi_N(i/N)=\begin{cases}-1/4, & i<\frac{N}{4}\\ \frac{i}{N}-\frac{1}{2}, & \frac{N}{4}\leqslant i\leqslant\frac{3}{4}N\\ 1/4, & i>\frac{3}{4}N\end{cases} \quad (2.7)$$

These scores are effective for detecting a shift in heavy-tailed distribution.

(4) Right-skewed distribution scores(denoted by RS) are

$$\phi_N(i/N)=\begin{cases}\frac{i}{N}-\frac{1}{2}, & i<\frac{N}{2}\\ 0, & i\geqslant\frac{N}{2}\end{cases} \quad (2.8)$$

They are effective for detecting shifts in distributions that are skewed to the right.

(5) Left-skewed distribution scores(denoted by LS) are

$$\phi_N(i/N)=\begin{cases}0 , & i<\frac{N}{2}\\ \frac{i}{N}-\frac{1}{2}, & i\geq\frac{N}{2}\end{cases} \tag{2.9}$$

They are effective for detecting shifts in distributions that are skewed to the left.

Evaluating the asymptotic relative efficiencies of the above tests using (2.4), we obtain Table 2.1. In the computations, the KW test was used as the bench mark for all of the other tests. The L test works well for the uniform, exponential and left-skewed beta distributions. For heavy-tailed distributions like double exponential and Cauchy, the H test is effective. The RS test is particularly effective for detecting shifts in the right-skewed exponential distribution. Table 2.1 also indicates that the LS test is good for detecting shifts in the left-skewed distributions.

In addition, Andrews(1954) shows that the Pitman efficiencies of the KW test relative to the classcal F test are 1, $3/\pi$, $\pi^2/9$, 1.5, and 3 for uniform, normal, logistic, double-exponential, and exponential distributions, respectively.

Table 2.1. Asymptotic Relative Efficiencies of Rank Tests

Distribution	ARE (L, KW)	ARE(H, KW)	ARE(RS, KW)	ARE(LS, KW)
Uniform	2.000	.500	.800	.800
Normal	.927	.870	.800	.800
Logistic	.781	.945	.800	.800
Double exponential	.500	1.125	.800	.800
Cauchy	.264	1.339	.800	.800
Exponential	2.000	.500	1.800	.200
Beta($\alpha=4$, $\beta=1$)	1.859	.536	.283	1.579

3. Adaptive Selection Rule

In this section we describe the use of rank-sum test statistics proposed in Section 2 to construct adaptive testing procedures that are truly nonparametric distribution-free. That is, the two stages of the inference process need be constructed in such a way that control of the overall significant α-level is maintained.

Consider again the hypothesis H_0 exposed in Section 2. If the underlying distribution $F(\cdot)$ is known, we can use rank sum test statistic associated with the score function that is effective for that case as is evidenced in Table 2.1 and 4.1. However, in reality we are not assuming the knowledge of distributional properties.

It would be desirable, therefore, to use the data itself first to select the model which seems most appropriate and then to perform the test with an appropriate set of scores based on the model chosen. On the procedures mentioned above the overall significant α-level of the test should be maintained.

Since each test statistic described in later part of Section 2 is distribution-free, it maintains the significance level α over the broad class of continuous distributions. Let $X_{(1)} \leqslant \ldots \leqslant X_{(N)}$ be the order statistics of combined sample of the $N=\sum_{i=1}^{k} n_i$ observations $X_{ij}(i=1, \ldots, k; j=1, \ldots, n_i)$. These order statistics are complete and sufficient for the underlying distribution $F(\cdot)$ under H_0. Thus, if we use a preliminary model selection scheme based on some statistic (or vector of statistics) computed from these order statistics the resulting adaptive test which comprises distribution-free rank tests is also distribution-free.

Note that the performances of the tests discussed in Section 2 depend on the tailweight and amount of skewness of underlying distributions. Thus, we might wish to select appropriate set of scores by using sample information about those population characteristics.

In this paper, we use a classification scheme originally proposed by Hogg, Fisher, and Randles (1975), which is based on skewness indicator Q_1 and kurtosis indicator Q_2, according to the papers of Uthoff(1970, 1973). These are

$$Q_1=(\bar{U}_{.05}-\bar{M}_{.50})/(\bar{M}_{.50}-\bar{L}_{.05}),$$

and (3.1)

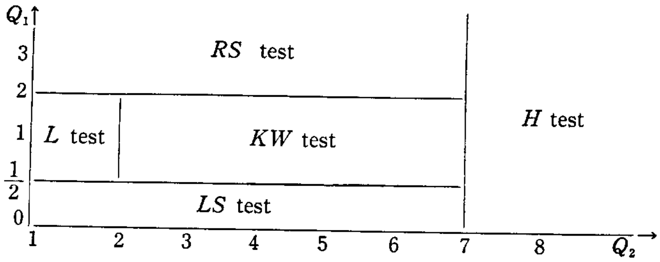
$$Q_2=(\bar{U}_{.05}-\bar{L}_{.05})/(\bar{U}_{.50}-\bar{L}_{.50}),$$

where $\bar{U}_r$($\bar{M}_r$ and $\bar{L}_r$) denotes the average of the rN largest(middle and smallest, respectively) combined order statistics $X_{(1)} \leqslant \ldots \leqslant X_{(N)}$. Fractional items are used when rN is not an integer; for example, if $N=30$, $\bar{U}_{.05}=(X_{(30)}+.5X_{(29)})/1.5$. When a light-tailed symmetric distribution is indicated ($.5 \leqslant Q_1 \leqslant 2$ and $Q_2<2$), we will use L test. If the data indicate medium-tailed or moderately heavy-tailed distribution ($.5 \leqslant Q_1 \leqslant 2$ and $2 \leqslant Q_2 \leqslant 7$), KW test is appropriate. The H test is used when a heavy-tailed distribution is indicated ($Q_2>7$). If the underlying distribution appears to be skewed to the right($Q_1>2$ and $Q_2 \leqslant 7$), then we will use RS test. If the data indicate the possibility that the underlying distribution is skewed to the left ($Q_1<.5$ and $Q_2 \leqslant 7$), LS test will be used. Figure 3.1 shows the classification and model selection scheme. The above procedure defines an adaptive distribution-free test for multi-sample location problem.

4. Monte Carlo Study on the Procedure

A Monte Carlo study was performed to investigate the performance of the adaptive distribution-free test discussed in Section 3. In the simulation study it was compared to both the classical F test and the component tests(KW, L, H, RS, and LS tests) of the adaptive rule.

Figure 3.1. Model Selection Scheme



The Monte Carlo study was run for the three-sample case with sample sizes $n_1=n_2=n_3=15$. The underlying distributions used were all members of a generalization of Tukey's lambda family, proposed by Ramberg and Schmeiser (1972, 1974). The inverse function of the generalized Tukey's lambda distribution(GLD) is given by

$$X=F^{-1}(P)=\lambda_1+(P^{\lambda_3}-(1-P)^{\lambda_4})/\lambda_2,\quad 0<P<1 \tag{4.1}$$

If P is a uniform(0, 1) random variable, then X has the GLD. The skewness and kurtosis of the GLD are determined by λ_3 and λ_4 (if $\lambda_3=\lambda_4$, it is symmetric), and given these values, the variance is determined by λ_2. The mean can then be shifted to any value by the appropriate choice of λ_1. The distributions used in this study are indexed by their skewness(α_3) and kurtosis(α_4) evaluated from the parameter values λ_3 and λ_4. The lambda family approximations to various distributions corresponding to the values of α_3 and α_4 are as follows.

	Lambda approximations to:									
	Uniform	Normal	Logistic	D-E	Heavy-tailed		Gamma (6, .5)	Exponential	Left-skewed	
α_3 :	0	0	0	0	0	0	816	.2.0	−.816	−2.0
α_4 :	1.8	3.0	4.19	6.0	13.2	126.9	4.0	9.0	4.0	9.0

FORTRAN program has been prepared for conducting the Monte Carlo study. The uniform random numbers were generated by RNDF function in CYBER 174/76. Fifteen uniform random numbers were generated and then inverse transformation was applied to generate the random sample using the inverse cdf. The process was repeated to generate the second and third sample. The transformed numbers for the second and third sample were incremented by different values of the shift parameter $\Delta_i (i=2, 3)$, and each of the seven tests was applied. All procedures were run at a nominal $\alpha=.05$ level. The cutoffs for the F test were chosen from the F distribution table, and those for the other rank sum tests were determined from their limiting chi-square distributions with 2 degrees of freedom.

The results of one such Monte Carlo study involving 5000 replications of the process for each of 11 distributions are shown in Table 4.1. In the table the adaptive distribution-free test is denoted by A. The first block ($\Delta_1=\Delta_2=\Delta_3=0$) in Table 4.1 presents the observed significance levels of the tests. The great discrepancy of these observed significance levels from .05 occurs in the F test for the Cauchy-like distribution. The observed significance levels of the LS test are slightly higher than the nominal level in all the underlying distributions under consideration.

The second block($\Delta_1=\Delta_2=0$, $\Delta_3=.6\sigma$) in Table 4.1 shows the powers of F, A, KW, L, H, RS, and LS tests when the third sample has been shifted by $\Delta_3=.6\sigma$, where σ denotes the standard deviation of the underlying distribution. The Cauchy-like distribution (having no moments) centered at zero was also derived from a generalized lambda distribution and was scaled so that $F^{-1}(.8413)=\sigma$, where .8413 is the probability below $\mu+\sigma$ in a normal distribution.

The third block($\Delta_1=0$, $\Delta_2=.6\sigma$, $\Delta_3=1.2\sigma$) in Table 4.1 indicates the powers of these tests under the alternative that the second and third sample have been shifted by $\Delta_2=.6\sigma$ and $\Delta_3=1.2\sigma$, respectively.

From the second and third block in Table 4.1, we note the power of the F test is extremely low when the underlying distribution is Cauchy-like. The adaptive distribution-free test A has a good power over F test in many instances. The KW test has also very good power properties. Except the light-tailed symmetric and most of the skewed distributions, the power of the KW test is slightly higher than but very close to that of the A test. The component tests which make up the adaptive one have quite remarkable power properties for the underlying distributions for which they are designed. We note that the results of the Monte Carlo study show the adaptive distribution-free tests have good power

Table 4.1. Empirical Powers Based on 5000 Replications with $n_1=n_2=n_3=15$

	Distribution										
α_3 :	0	0	0	0	0	0	Cauchy-	.816	2.0	−.816	−2.0
α_4 :	1.8	3.0	4.19	6.0	13.2	126.9	like	4.0	9.0	4.0	9.0
Test	$\Delta_1=\Delta_2=\Delta_3=0$										
F	.045	.047	.052	.052	.048	.047	.015	.047	.043	.048	.046
A	.046	.046	.048	.051	.055	.053	.048	.047	.042	.050	.066
KW	.046	.046	.049	.051	.055	.053	.050	.047	.051	.050	.057
L	.044	.048	.054	.049	.057	.049	.047	.047	.046	.049	.048
H	.043	.043	.049	.050	.052	.053	.052	.048	.053	.049	.060
RS	.035	.042	.040	.041	.046	.045	.043	.044	.043	.041	.046
LS	.059	.055	.063	.062	.060	.061	.062	.054	.063	.057	.066
Test	$\Delta_1=\Delta_2=0,\ \Delta_3=.6\sigma$										
F	.337	.355*	.350	.379	.382	.425	.043	.369	.380	.350	.395
A	(.384)	(.332)	(.371)	(.425)	(.465)	(.563)	(.366)	(.389)	(.804)	(.401)	(.769)
KW	.310	.334	.374*	.429*	.473	.576	.349	.402*	.633	.384	.597
L	.489*	.303	.296	.309	.327	.370	.129	.326	.409	.396	.757
H	.210	.301	.354	.425	.485*	.589*	.445*	.375	.628	.327	.454
RS	.212	.217	.230	.259	.295	.353	.218	.356	.819*	.173	.168
LS	.289	.332	.371	.425	.471	.578	.381	.295	.322	.475*	.801*
Test	$\Delta_1=0,\ \Delta_2=.6\sigma,\ \Delta_3=1.2\sigma$										
F	.822	.813*	.804	.813	.824	.843	.111	.822	.819	.816	.818
A	(.770)	(.790)	(.825)	(.861)	(.903)	(.947)	(.716)	(.852)	(.969)	(.849)	(.972)
KW	.763	.793	.829*	.868*	.912	.953	.701	.853	.945	.849	.948
L	.917*	.730	.694	.713	.737	.787	.278	.770	.877	.798	.915
H	.574	.743	.808	.858	.916*	.961*	.814*	.822	.939	.798	.931
RS	.606	.639	.687	.736	.798	.865	.561	.859*	.993*	.568	.564
LS	.657	.718	.753	.794	.856	.911	.655	.642	.645	.882*	.993*

The values in the parentheses are the powers are the powers of adaptive distribution-free tests(A).
* indicates the highest power in each case.

properties over a wide class of underlying distributions.

5. Discussion

In the process constructing the adaptive model selection scheme we consider only 5 rank tests as the components of the adaptive rule, but it would be possible to increase the number of rank tests used in the adaptive scheme. However, it is not worthwhile to create additional tests unless a substantial increase in power is attainable with them.

The results of the Monte Carlo study indicate the adaptive distribution-free tests are very successful when the underlying distribution is unknown.

Finally, it is mentioned that the adaptive procedure would be also favourable in statistical subjects like two-way analysis of variance problem, slippage problem, and etc., and are certainly worth considering in practical problems.

REFERENCES

(1) Andrews, F.C. (1954). Asymptotic Behavior of Some Rank Tests for Analysis of Variance, *Ann. Math. Statist.*, Vol. 25, 724—36.

(2) Hâjek, J. and Sidâk, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

(3) Hogg, R.V., Fisher, D.M. and Randles, R.H. (1975). A Two-sample Adaptive Distribution-free Test, *J. Amer. Statist. Assoc.*, Vol.70, 656—61.

(4) Kruskal, W.H. and Wallis, W.A. (1952). Use of Ranks in One Criterion Analysis of Variance, *J. Amer. Statist. Assoc.*, Vol.47, 583—621.

(5) Puri, M.L. (1964). Asymptotic Efficiency of a Class of C-sample Tests, *Ann. Math. Statist.*, Vol.35, 102—21.

(6) Ramberg, J.S. and Schmeiser, B.W. (1972). An Approximate Method for Generating Symmetric Random Variables, *Comm. of the A.C.M.*, Vol. 15, 987—90.

(7) Ramberg, J.S. and Schmeiser, B.W. (1974). An Approximate Method for Generating Asymmetric Random Variables, *Comm. of the A.C.M.*, Vol. 17, 78—82.

(8) Randles, R.H. and Hogg, R.V. (1973). Adaptive Distribution-Free Tests, *Comm. in Statist.*, Vol. 2(4), 337—56.

(9) Randles, R.H. and Wolfe, D.A. (1979). *Introduction to the Theory of Nonparametric Statistics*, John Wiley & Sons, Inc., New York.

(10) Uthoff, V.A. (1970). An Optimum Test Property of Two Well Known Statistics, *J. Amer. Statist. Assoc.*, Vol. 65, 1597—1600.

(11) Uthoff, V.A. (1973). The Most Powerful Scale and Location Invariant Test of the Normal Versus Double Exponential, *Ann. Statist.*, Vol. 1, 170—4.