*Appl. Statist.* (1988)
**37**, *No.* 2, *pp.* 205–218

# Adaptive Nonparametric Procedures and Applications

By N. J. HILL

*Methodist Ladies College, Kew, Australia*

A. R. PADMANABHAN†

*Monash University, Clayton, Australia*

MADAN L. PURI

*Indiana University, Bloomington, USA*

[Received January 1985. Final revision December 1987]

SUMMARY

Two adaptive nonparametric procedures are proposed for multiple comparisons and testing for ordered alternatives in the one-way ANOVA model. The first procedure resembles a proposal of Hogg, Fisher and Randles (for hypothesis testing) while the second is a variation of the first. Applications to data on lung cancer illustrate the theory.The supremacy of these procedures over the parametric normal theory procedures and the rank-based procedures is established. Monte Carlo studies show that these procedures can be safely applied when the size of each sample is at least 20.

*Keywords*: Multiple comparisons; Ordered alternatives; One-way ANOVA; Lung cancer data

## 1. Introduction

This paper develops adaptive inference for multiple comparisons and tests of ordered alternatives in the one-way analysis of variance (ANOVA) model. It has been motivated by the following two considerations.

(a) For hypothesis testing in the two-sample location model, the work of Hogg *et al.* (1975) has shown the supremacy of adaptive procedures over (i) the usual nonparametric procedure of working only with ranks and the resulting statistics such as Wilcoxon and (ii) the parametric normal theory procedure.
(b) In addition, the results of Puri (1964, 1965), Puri and Sen (1971) and Hajek and Sidak (1967) ensure that these good properties of the adaptive procedures extend to multiple comparisons and testing for ordered alternatives in the one-way ANOVA model. Consequently, we can have adaptive inference in the one-way ANOVA model as well.

Section 2 discusses the adaptive procedures. Section 3 contains testing for unrestricted alternatives. Sections 4 and 5 describe respectively multiple comparisons and testing for ordered alternatives. Section 6 discusses applications, while Section 7 is devoted to some theoretical results and Monte Carlo studies. Section 8 summarises the main findings.

† *Address for correspondence*: Department of Mathematics, Monash University, Clayton, Victoria 3168, Australia.

 0035–9254/88/37205 $2.00

## 2. Adaptive Procedures

Some two-sample statistics (on which the adaptive procedures are based) will be defined first.

Let $X_1 = (X_{11}, \ldots, X_{1n_1})$ and $X_2 = (X_{21}, \ldots, X_{2n_2})$ be two random samples from respective continuous distribution functions $F_1(x) = F(x - \theta_1)$ and $F_2(x) = F(x - \theta_2)$. Write $\Delta = \theta_1 - \theta_2$. Let $R_{1i}$ be the rank of $X_{1i}$ in the combined sample $(X_1, X_2)$ and $a(1), a(2), \ldots, a(n_1 + n_2)$ be a set of non-decreasing scores with $a(1) \neq a(n_1 + n_2)$. Tests of the hypothesis $H_0$: $\Delta = 0$ against the alternative $H_1$: $\Delta > 0$ are usually based on rank statistics of the form

$$h = h(X_1, X_2) = \sum_{i=1}^{n_1} a(R_{1i}). \tag{2.1}$$

The scores $a_{\mathrm{L}}(i)$, $a_{\mathrm{ML}}(i)$, $a_{\mathrm{W}}(i)$, $a_{\mathrm{SR}}(i)$ and $a_{\mathrm{SL}}(i)$ will now be defined. When substituted in the right-hand side of equation (2.1), these yield the statistics, say, $h_{\mathrm{L}}$, $h_{\mathrm{ML}}$, $h_{\mathrm{W}}$, $h_{\mathrm{SR}}$ and $h_{\mathrm{SL}}$ respectively.

For any positive number $B$, let $[B]$ denote the largest integer less than or equal to $B$. Then $a_{\mathrm{L}}(i) = a_{\mathrm{L}}^*(i)/(n_1 + n_2 + 1)$, where

$$a_{\mathrm{L}}^*(i) = \begin{cases} i - \left(\dfrac{n_1 + n_2 + 1}{4}\right) - \dfrac{1}{2}, & i \leqslant \dfrac{n_1 + n_2 + 1}{4} \\ i - (n_1 + n_2) + \dfrac{n_1 + n_2 + 1}{4} - \dfrac{1}{2}, \text{ if } i \geqslant n_1 + n_2 - \left(\dfrac{n_1 + n_2 + 1}{4}\right) + 1 \\ 0, \text{ otherwise.} \end{cases} \tag{2.2}$$

$$a_{\mathrm{ML}}(i) = \frac{a_{\mathrm{ML}}^*(i)}{(n_1 + n_2 + 1)^2},$$

where

$$a_{\mathrm{ML}}^*(i) = \begin{cases} -\left[i - \left(\dfrac{n_1 + n_2 + 1}{4}\right) - \dfrac{1}{2}\right]^2, \text{ if } i \leqslant \dfrac{n_1 + n_2 + 1}{4} \\ \left[i - (n_1 + n_2) + \dfrac{n_1 + n_2 + 1}{4} - \dfrac{1}{2}\right]^2, \text{ if } i \geqslant n_1 + n_2 - \left(\dfrac{n_1 + n_2 + 1}{4}\right) + 1 \\ 0, \text{ otherwise} \end{cases} \tag{2.3}$$

$$a_{\mathrm{W}}(i) = \frac{a_{\mathrm{W}}^*(i)}{n_1 + n_2 + 1}, \text{ where } a_{\mathrm{W}}^*(i) = i,\ 1 \leqslant i \leqslant n_1 + n_2 \tag{2.4}$$

(i.e. the Wilcoxon scores).

$$a_{\mathrm{SR}}(i) = \frac{a_{\mathrm{SR}}^*(i)}{n_1 + n_2 + 1}$$

where

$$a_{\mathrm{SR}}^*(i) = \begin{cases} i - \left(\dfrac{n_1 + n_2 + 1}{2}\right) - 1, \text{ if } i \leqslant \dfrac{n_1 + n_2 + 1}{2} \\ 0, \text{ otherwise.} \end{cases} \tag{2.5}$$

$$a_{\mathrm{SL}}(i) = \frac{a_{\mathrm{SL}}^*(i)}{n_1 + n_2 + 1},$$

where

$$a^*_{\mathrm{SL}}(i) = \begin{cases} i - \left(\dfrac{n_1+n_2+1}{2}\right), & \text{if } i \geqslant \dfrac{n_1+n_2+1}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

In many situations $F$ is

(a) light tailed and symmetric or
(b) skewed.

Examples of (a) are discussed by Hogg, Box and Cox (see Hogg (1974), p. 914, Harter (1974) and Wegman and Carroll (1977), p. 796) and are also found in some Australian high school examinations scores. Some examples of (b) are given in Section 5. In general $F$ is unknown. Nevertheless, suppose there is a preliminary classification which detects the tail weight and skewness of $F$. Then the rank test suggested by this scheme will surpass the Wilcoxon and $t$ tests. As this supremacy also extends to the $c$ sample problem, such an adaptive scheme will have far reaching consequences.

One such scheme (Hogg, 1974) uses as a measure of tail weight

$$\kappa = (E_{0.05} - G_{0.05})/(E_{0.5} - G_{0.5}),$$

where $E_\alpha$ is the mean of the upper $100\alpha\%$ truncated distribution and $G_\alpha$ is the mean of the lower $100\alpha\%$ truncated distribution, $\alpha = 0.05$, 0.5 (provided that these means exist). $\kappa$ will be estimated by

$$Q_2 = (\bar{U}_{0.05} - \bar{L}_{0.05})/(\bar{U}_{0.5} - \bar{L}_{0.5}),$$

where $\bar{U}_\alpha$ ($\bar{L}_\alpha$) is the mean of the upper (lower) $100\alpha\%$ order statistics of the combined sample ($X_1$, $X_2$). $Q_2$ has been shown by extensive Monte Carlo studies to be a good indicator of tail weight when $\Delta$ is close to zero. $Q_2$ may indicate the wrong (test) statistic if the shift $\Delta$ is large. Most rank tests detect a large shift with a high probability, so this is not a serious problem from the point of view of testing. However, an inappropriate statistic may yield a long confidence interval, which is a serious problem from the point of view of estimation. Therefore we work with

$$\bar{Q}_2 = (n_1 Q_{2,1} + n_2 Q_{2,2})/(n_1 + n_2),$$

the weighted average of the $Q_2$ values based on the individual samples. As $\bar{Q}_2$ is unaffected by the actual value of $\Delta$, it outperforms $Q_2$ in detecting the tail weight.

An indicator of skewness, studied by Fisher and explained in Hogg *et al.* (1975), is $Q_1 = (\bar{U}_{0.05} - \bar{M}_{0.5})/(\bar{M}_{0.5} - \bar{L}_{0.5})$, $\bar{M}_{0.5}$ being the average of the middle 50% of the order statistics of the combined sample. However, for reasons explained earlier, we work with

$$\bar{Q}_1 = (n_1 Q_{1,1} + n_2 Q_{1,2})/(n_1 + n_2),$$

the weighted average of the $Q_1$ values based on the individual samples. Following Hogg *et al.* (1975), we can say that the assumption of symmetry is tenable if $\frac{1}{2} \leqslant \bar{Q}_1 \leqslant 2$. ($\bar{Q}_1 > 2$ and $\bar{Q}_1 < \frac{1}{2}$ indicate skewness to the right and left respectively.)

The $\kappa$ values for the rectangular and normal distributions are respectively 1.9 and 2.58. The midpoint of (1.9, 2.58) is 2.24. Suppose now $\bar{Q}_1$ suggests symmetry. Let $\bar{Q}_2 < 2.24$. This implies that $F$ is closer to the rectangular distribution (than the normal distribution), and hence can be classified as light tailed (or L), whereas if $\bar{Q}_2 \geqslant 2.24$ $F$ has to be classified as not light tailed (or NL).

On the basis of theoretical, empirical and intuitive considerations, $F$ is classified as heavy tailed (or H) if $\bar{Q}_2 \geqslant 3.8$ and not heavy tailed (or NH) if $\bar{Q}_2 < 3.8$. This classification applies regardless of whether $F$ is symmetric or skewed.

The light-tailed symmetric case requires scores that emphasise the extreme observations. We then use $h_{\mathrm{L}}$ or $h_{\mathrm{ML}}$. (The asymptotic relative efficiencies (AREs) of the $h_{\mathrm{ML}}$ test and the $h_{\mathrm{L}}$ test are respectively 3.33 and 2 in the rectangular case, and 0.8 and 0.88 in the normal case. Hence they should be effective for the category L.)

The case where $F$ is classified as NH and right skewed requires scores that emphasise the smallest observations. We then use $h_{\mathrm{SR}}$. Similarly $h_{\mathrm{SL}}$, whose scores emphasise the largest observations, is appropriate for the data which is NH and left skewed. When we simply say 'skewed' we mean 'right skewed'.

We now describe the two adaptive schemes. The first resembles that of Hogg *et al.* (1975), while the second is a variation of the first.

## 2.1. *Schemes I and II*

Suppose we have prior knowledge that the data are skewed and NH (there are many such instances, some of which are mentioned in Section 5). Then we work with $h_{\mathrm{SR}}$ straightaway, whereas, if the data are known to be left skewed and NH, we work with $h_{\mathrm{SL}}$. In both these cases, *it is not necessary to calculate* $\bar{Q}_1$ and $\bar{Q}_2$. The remaining cases are covered in Table 1.

The extension of these considerations to the $c$ sample case is straightforward. Now $\bar{Q}_i$, $i = 1, 2, \ldots, c$, will denote the average of the $Q_i$ values computed for each of the $c$ samples.

The definitions of the statistics in this section are based on the assumption that the observations are all distinct. However, in practice frequently two or more observations take the same value, resulting in a tie. We now show that by using the method of average scores the foregoing statistics can be defined even in the presence of ties.

Consider the two-sample statistic $h = h(X_1, X_2)$ given by equation (2.1) based on the scores $a(1), a(2), \ldots, a(n_1 + n_2)$. Let $Z_1 \leqslant Z_2 \leqslant, \ldots, Z_{n_1+n_2}$ be the order statistics of the combined sample. Suppose we have

$$Z_{r_1} < Z_{r_1+1} = Z_{r_1+2} = \ldots = Z_{r_1+K} < Z_{r_1+K+1}.$$

*Case 1.* $r_1 \geqslant 1$. Then the observations $Z_{r_1+1}, \ldots, Z_{r_1+K}$ tie. If they had been distinct, they would have among themselves accounted for the scores $a(r_1 + 1)$,

TABLE 1
*Adaptive schemes when we have no prior knowledge*

| *Indicator values* | *For adaptive scheme I, use* | *For adaptive scheme II, use* |
|---|---|---|
| $Q_2 \geqslant 3.8$ | $h_{\mathrm{W}}$ | $h_{\mathrm{W}}$ |
| $\frac{1}{2} \leqslant \bar{Q}_1 \leqslant 2$ and $2.24 \leqslant \bar{Q}_2 < 3.8$ | $h_{\mathrm{W}}$ | $h_{\mathrm{W}}$ |
| $\frac{1}{2} \leqslant \bar{Q}_1 \leqslant 2$ and $\bar{Q}_2 < 2.24$ | $h_{\mathrm{L}}$ | $h_{\mathrm{MJ}}$ |
| $\bar{Q}_1 < \frac{1}{2}$ and $\bar{Q}_2 < 3.8$ | $h_{\mathrm{SL}}$ | $h_{\mathrm{SL}}$ |
| $\bar{Q}_1 > 2$ and $\bar{Q}_2 < 3.8$ | $h_{\mathrm{SR}}$ | $h_{\mathrm{SR}}$ |

$a(r_1+2), \ldots, a(r_1+K)$. The resulting sum of scores would have been $a(r_1+1)+a(r_1+2)+\ldots+a(r_1+K)$. Distribute this sum equitably over all of them, i.e. assign to each of them the average score $(1/K)[a(r_1+1)+\ldots+a(r_1+K)]$.

*Case 2.* $r_1=0$ and we have the tie $Z_1=Z_2=\ldots=Z_k<Z_{K+1}$. Now each member in this tie is assigned the score $(1/K)[a(1)+\ldots+a(K)]$.

Cases 1 and 2 together cover all possible ties.

Let $\tilde{a}(1), \tilde{a}(2), \ldots, \tilde{a}(n_1+n_2)$ be the new scores, resulting from the arrangement in cases 1 and 2. Then we define

$$h(X_1, X_2)=\sum_{i=1}^{n_1} \tilde{a}(R_{1i}).$$

Using this method, the $c$ sample statistics of Sections 3 and 5 can be similarly defined in the presence of ties. For further details see Hajek (1969), p. 129.

## 3. Testing for Unrestricted Alternatives

Let there be $c$ independent samples, where the $i$th sample, say $X_i$ $(i=1, 2, \ldots, c)$, comprises $n_i$ independent and identically distributed random variables $X_{i1}, X_{i2}, \ldots, X_{in_i}$ having a common absolutely continuous distribution function $F_i(x)$. Thus $X_i=(X_{i1}, X_{i2}, \ldots, X_{in_i})$.

Let $N=n_1+n_2+\ldots+n_c$ and $Y_1 \leqslant Y_2 \leqslant \ldots \leqslant Y_N$ be the $N$ elements $X_{ij}$ $(j=1, 2, \ldots, n_i;\ i=1, 2, \ldots, c)$ arranged in ascending order. Let $a(1), a(2), \ldots, a(N)$ be a set of scores, whose average is $\bar{a}_N$.

Replace $Y_1, Y_2, \ldots., Y_N$ by $a(1), a(2), \ldots, a(N)$ respectively, and denote by $S_i$ the resulting sum of scores for the $i$th sample.

We now make the following two assumptions.

(a) $F_i(x)=F(x-\theta_i)$, $i=1, 2, \ldots, c$.
(b) The scores $a(1), a(2), \ldots, a(N)$ satisfy some mild regularity conditions (see Puri (1964)).

To test $H_0$: $\theta_1=\theta_2=\ldots=\theta_c$ against $H_1$: $\theta_r \neq \theta_s$ for at least one pair $(r, s)$, use

$$S_c=(N-1)\sum_{i=1}^{c} n_i\left(\frac{S_i}{n_i}-\bar{a}_N\right)^2 \bigg/ \sum_{r=1}^{N} (a(r)-\bar{a}_N)^2. \tag{3.1}$$

If ties occur, they are handled by the average scores method. Then theorem 4.5 of Conover (1973) implies that, under $H_0$, $S_c$ has asymptotically a chi-square distribution with $c-1$ degrees of freedom. Let its $(1-\alpha)$th quantile be denoted by $\chi^2_{1-\alpha}$ (for some pre-assigned level of significance $\alpha$).

If the value of $S_c$ computed from the sample exceeds $\chi^2_{1-\alpha}$, then $H_0$ should be rejected. Consequently, the study of multiple comparisons becomes relevant and will be pursued in Section 4.

## 4. Multiple Comparisons

Once the appropriate scores have been chosen, the simultaneous confidence intervals are simply those given on pp. 247, 250 and 254 of Puri and Sen (1971) and can be computed as in Bauer (1972). The fact that they have the required (asymptotic) coverage probabilities follows from the arguments in Puri and Sen, and the observations that the procedures are asymptotically distribution free (cf. Section 7). Moreover, ties, when handled by the average scores method, have little effect on these intervals (see Lehmann (1975) and Padmanabhan (1977)).

Some of these intervals involve the $(1-\alpha)$th quantile of the range of a sample of size $c$ from a standard normal distribution. Such values are given in Harter (1960).

## 5. Testing for Ordered Alternatives

In the notation of Section 3, consider the problem of testing $H_0$: $\theta_1 = \theta_2 = \ldots = \theta_c$ against the ordered alternative $H_A$: $\theta_1 \leqslant \ldots \leqslant \theta_c$ (or $H_A$: $\theta_1 \geqslant \theta_2 \geqslant \ldots \geqslant \theta_c$), where at least one of the inequalities is strict. Probably the best-known nonparametric solution is the Jonckheere test, whose validity requires no knowledge of $F$, apart from its continuity. However, there are many situations where $F$ is known to be skewed and not heavy tailed, although its actual form may be unknown. Some such examples are

(a) lifetimes of cancer patients;
(b) distribution of radii or aerosols;
(c) distribution of age at death of infants, dying from respiratory distress syndrome;
(d) biomass contained in a unit volume of water;
(e) the toxicity of some drugs, food additives etc;
(f) many problems in life testing (see Chen (1982)).

A test will be constructed which is more powerful than the Jonckheere test in the foregoing situations.

For the problem under study, Puri (1965) proposed a family of $c$ sample statistics, say $V$, which is an adaptation of a two-sample statistic $h$ to the $c$ sample situation. Theorem 5.3 of Puri (1965) implies that, if the test based on $h$ is good for the two-sample problem, then the test based on $V$ is good for the $c$ sample problem. More precisely, let $h_1$ and $h_2$ be two-sample statistics and $V_1$ and $V_2$ the corresponding $V$ statistics. Then the asymptotic efficiency of $V_1$ relative to $V_2$ is the same as the asymptotic efficiency of $h_1$ relative to $h_2$. We shall exploit this result and construct $V$ based on the appropriate scores.

We begin by defining a $V$ statistic based on general scores and show how the Jonckheere statistic follows as a special case. We then go on to construct a statistic tailored for skewed data and finally obtain its asymptotic null distribution.

Choose two arbitrary samples $X_i$ (of size $n_i$) and $X_j$ (of size $n_j$). For definiteness, let $i<j$. Suppose that $a(1), a(2), \ldots, a(n_i+n_j)$ are the scores corresponding to $h(X_i, X_j)$. (See formulae (2.2)–(2.6)). Let $R_{ik}$ denote the rank of $X_{ik}$ in the combined sample $(X_i, X_j)$. Then we replace $X_{ik}$ by $a(R_{ik})$. Thus

$$h(X_i, X_j) = \sum_{k=1}^{n_i} a(R_{ik}).$$

Write $A_{ij} = a(1) + a(2) + \ldots + a(n_i + n_k)$. Then the corresponding $V$ statistic (see Puri (1965)) is defined by

$$V = \sum\sum_{1 \leqslant i < j \leqslant c} [(n_i + n_j)h(X_i, X_j) - n_i A_{ij}] \ldots \tag{5.1}$$

Let $V_{\mathrm{W}}$ and $V_{\mathrm{SR}}$ be respectively the $V$ statistics corresponding to the choices $h = h_{\mathrm{W}}$

(the Wilcoxon statistic) and $h = h_{SR}$ based on the scores $a_{SR}$ (see formula (2.5)). Then $V_W$ is equivalent to the Jonckheere statistic (see Puri (1965)).

*Remark.* It appears as if our definition of the $V$ statistic is somewhat different from that of Puri (cf. formula (2.4) of Puri (1965)). However, it can be shown that they are really identical. Moreover, our definition is preferable from the computational point of view.

Write $N = n_1 + n_2 + \ldots + n_c$ and $\rho_i = n_i/N$. It is well known (see Puri (1965)) that, under $H_0$, $N^{-3/2} V_{SR}$ is asymptotically normal with mean zero and variance

$$\frac{5}{576}(1 - \Sigma\rho_i^3). \tag{5.2}$$

When many ties occur (see examples 1 and 2 of Section 6), the variance given by formula (5.2) is not reliable. Therefore we give an alternative expression. Set

$$a_N(r) = \frac{1}{N}\left[(r-1) - \left(\frac{N+1}{2}\right)\right], \text{ for } r \leqslant \frac{N+1}{2}$$
$$= 0, \text{ for } r > \frac{N+1}{2}.$$

Write $\bar{a}_N = (1/N)\{a_N(1) + a_N(2) + \ldots + a_N(N)\}$. Let $\tilde{a}_N(\cdot)$ be obtained from $a_N(\cdot)$, when ties (in the combined sample of $c$ groups) are handled by the average scores method. Let

$$\tilde{A}^2 = \frac{1}{N-1}\{\Sigma(\tilde{a}_N(r) - \bar{a}_N)^2\}$$

and

$$\tilde{\sigma}^2 = \frac{1}{3}[1 - (\Sigma\rho_i^3)]\tilde{A}^2. \tag{5.3}$$

Theorem 4 of Vorlickova (1970) and arguments similar to theorem 5.2 of Puri (1965) imply that, in the presence of ties, the asymptotic null distribution of $N^{-3/2} V_{SR}$ is normal, with mean zero and variance $\tilde{\sigma}^2$.

For data which are skewed and not heavy tailed, the $V_{SR}$ test is superior to the $V_W$ test (see Section 7, part I).

## 6. Applications

Tables 2 and 3 contain data on survival days of patients with inoperable lung cancer, who were subjected to a standard chemotherapeutic agent (standard, for short) and a test chemotherapeutic agent (test, for short) respectively. Within each table, patients were divided, depending on the histological type of tumour, into the following categories: squamous, small, adeno and large.

These data are part of the data, which were collected by the Veterans Administrative Lung Cancer Study Group, in a multi-institutional collaborative trial in the USA.

The resulting distributions are skewed and not heavy tailed and the assumption of ordered alternative is tenable.

The groups (in the four categories) may be assumed to be the samples from the distributions $F(x - \theta_1)$, $F(x - \theta_2)$, $F(x - \theta_3)$ and $F(x - \theta_4)$ respectively. Since $c > 2$ and

TABLE 2

*Survival days of patients (standard chemotherapy)*

| | |
|---|---|
| Group 1 (standard, squamous) | 72, 411, 228, 126, 118, 10, 81, 110, 314, 100, 42, 8, 144, 25, 11 |
| Group 2 (standard, small) | 30, 384, 4, 54, 13, 23, 97, 153, 59, 117, 16, 151, 22, 56, 21, 18, 139, 20, 31, 52, 287, 18, 51, 122, 27, 54, 7, 63, 392, 10 |
| Group 3 (standard, adeno) | 8, 92, 35, 117, 132, 12, 162, 3, 95 |
| Group 4 (standard, large) | 177, 162, 216, 553, 278, 12, 260, 200, 156, 182, 143, 105, 103, 250, 100 |

TABLE 3

*Survival days of patients (test chemotherapy)*

| | |
|---|---|
| Group 1 (test, squamous) | 999, 112, 242, 991, 111, 1, 587, 389, 38, 25, 357, 467, 201, 1, 30, 44, 283, 15 |
| Group 2 (test, small) | 25, 21, 13, 87, 2, 20, 7, 24, 99, 8, 99, 61, 25, 95, 80, 52, 29 |
| Group 3 (test, adeno) | 24, 18, 31, 51, 90, 52, 73, 8, 36, 48, 7, 140, 186, 84, 19, 45, 80 |
| Group 4 (test, large) | 52, 164, 19, 53, 15, 43, 340, 133, 111, 231, 378, 49 |

each sample size exceeds five, the approximation provided by the asymptotic theory of $V$ is adequate (see Lehmann (1975), p. 207). An alternative $\theta_1 \leqslant \theta_2 \leqslant \theta_3 \leqslant \theta_4$, with at least one strict inequality, points to negative values of $N^{-3/2}V_{SR}$, so that the critical region will be the appropriate left-hand tail of the limiting normal distribution. Similarly, for the alternative $\theta_1 \geqslant \theta_2 \geqslant \theta_3 \geqslant \theta_4$, with at least one strict inequality, the critical region will be the corresponding right-hand tail.

However, although a monotonic trend exists, its direction is not sure, i.e. whether $\theta_1 \geqslant \theta_2 \geqslant \theta_3 \geqslant \theta_4$ or $\theta_1 \leqslant \theta_2 \leqslant \theta_3 \leqslant \theta_4$. We shall therefore test $H_0$ against the two-sided alternative $\theta_1 \geqslant \theta_2 \geqslant \theta_3 \geqslant \theta_4$ or $\theta_1 \leqslant \theta_2 \leqslant \theta_3 \leqslant \theta_4$ (with at least one strict inequality in each case).

Now $V_{SR}$ is a combination of six two-sample statistics $h(X_1, X_2)$, $h(X_1, X_3)$, $h(X_1, X_4)$, $h(X_2, X_3)$, $h(X_2, X_4)$ and $h(X_3, X_4)$ (all of which are based on the scores $a_{SR}$ (see formula (2.5)). As ties abound, all these statistics have to be defined using the average scores method. After performing the computations, we find that for the data $V_{SR} = -99.46$, $N = 69$, $N^{-3/2}V_{SR} = -0.17$ and $\tilde{\sigma}^2 = 0.008$. The upper 0.025 quantile of the normal distribution with zero mean and variance $\tilde{\sigma}^2$ is $0.18 = \lambda$ (say). Hence the two-sided critical region is $(-\infty, -\lambda) \cup (\lambda, \infty)$. As $N^{-3/2}V_{SR}$ is outside this region. $H_0$ is accepted. In fact, the significance probability is $2 \times 0.029 = 0.058$.

Under the same notation and assumptions as in the analysis of Table 2 we have in this case $N^{-3/2}V_{SR} = -0.0019$ and $\tilde{\sigma}^2 = 0.008$, so that $\tilde{\sigma} = 0.09$. Since the upper 0.025 quantile of the normal distribution with mean zero and standard deviation $\tilde{\sigma}$ is approximately $0.18 = \lambda$ (say), the two-sided critical region is $(-\infty, -\lambda) \cup (\lambda, \infty)$. As $N^{-3/2}V_{SR}$ is well outside this critical region, $H_0$ is accepted.

## 7. Some Theoretical Results and Monte Carlo Studies

### 7.1. *Part I*

At the end of Section 5, it was stated that for data, which are skewed and not heavy tailed, the $V_{SR}$ test is superior to the $V_W$ test. To establish this result, we evaluate ARE($V_{SR}$, $V_W$), the asymptotic efficiency of the $V_{SR}$ test relative to the $V_W$ test, for two important biomedical distributions, namely the exponential and log-normal distributions.

The score functions $V_{SR}$ and $V_W$ are respectively $J_{SR}$ and $J_W$ defined by

$$\begin{aligned} J_{SR}(u) &= u - \tfrac{1}{2}, \quad 0 < u < \tfrac{1}{2} \\ &= 0, \text{ elsewhere} \end{aligned} \tag{7.1}$$

$$\begin{aligned} J_W(u) &= u, \quad 0 < u < 1 \\ &= 0, \text{ elsewhere} \end{aligned} \tag{7.2}$$

(see Randles and Wolfe (1979)).

From Puri (1965), it follows that

$$\text{ARE}(V_{SR}, V_W) = \frac{A_W^2 B_{SR}^2}{A_{SR}^2 B_W^2} \tag{7.3}$$

where

$$A_W^2 = \int_0^1 J_W^2(u)\, du - \left(\int_0^1 J_W(u)\, du\right)^2 \tag{7.4}$$

and

$$B_W = \int_{-\infty}^{\infty} [dJ_W(F(x))/dx]\, dF(x). \tag{7:5}$$

(Here $F$ is the underlying distribution.) $A_{SR}^2$ and $B_{SR}$ are obtained by replacing $J_W$ by $J_{SR}$ in equations (7.4) and (7.5) respectively. Clearly

$$A_W^2 = \frac{1}{12} \text{ and } A_{SR}^2 = \frac{5}{192}. \tag{7.6}$$

7.1.1. *Case 1: Exponential Distribution.* Now $F(x) = 1 - \exp(-x)$ and its density is

$$\begin{aligned} f(x) &= \exp(-x), \quad x \geqslant 0 \\ &= 0, \text{ otherwise.} \end{aligned}$$

Hence

$$B_{SR} = \int_0^{\ln 2} f^2(x)\, dx = \frac{3}{8} \tag{7.7}$$

and

$$B_W = \int_0^{\infty} f^2(x)\, dx = \frac{1}{2}. \tag{7.8}$$

Substituting equations (7.6)–(7.8) in equation (7.2), we obtain $\mathrm{ARE}(V_{SR}, V_W) = 1.8$. As this $\mathrm{ARE} > 1$, the $V_{SR}$ test is now superior to the $V_W$ test.

7.1.2. *Case 2: Log-normal Distribution.* The density is

$$f(x) = \frac{1}{x\sqrt{(2\pi)}\,t} \exp\left[-\frac{1}{2}\frac{(\ln x - m)^2}{t^2}\right], \qquad x > 0,\ t > 0.$$

Routine computation shows that

$$B_{SR} = \frac{\exp(t^2/4)\Phi(t\sqrt{2})}{2t\sqrt{\pi}\exp m}$$

(where $\Phi(\cdot)$ is the standard normal distribution function) and

$$B_W = \frac{\exp(t^2/4)}{2t\sqrt{\pi}\exp m}.$$

Hence

$$\mathrm{ARE}(V_{SR}, V_W) = \frac{16}{5}\,[\Phi(t/\sqrt{2})]^2.$$

Since $t > 0$, it is easy to verify that

$$0.80 < \mathrm{ARE}(V_{SR}, V_W) < 3.2 \tag{7.9}$$

and

$$\mathrm{ARE} > 1, \text{ as soon as } t > 0.2017. \tag{7.10}$$

$(0 < t < \infty)$. Equations (7.9) and (7.10) show now that the $V_{SR}$ test is almost always superior to the $V_W$ test and occasionally only slightly inferior.

### 7.2. *Part II*

Propositions 7.1 and 7.2 establish that our adaptive procedures are asymptotically distribution free. Monte Carlo studies show that, even for reasonable sample sizes,

(a) the foregoing distribution-free property becomes meaningful,
(b) the adaptive tests have generally higher power than the Wilcoxon and $t$ test and
(c) the adaptive confidence intervals are generally shorter than those given by the Wilcoxon or $t$ tests.

*Proposition 7.1.* Suppose that the adaptive scheme is based on $(Q_1, Q_2)$ (defined in Section 2). Then it is exactly distribution free. The proof is well known (see Randles and Hogg (1973)).

*Proposition 7.2.* Under $H_0$, the schemes based on $(Q_1, Q_2)$ and $(\bar{Q}_1, \bar{Q}_2)$ are asymptotically equivalent.

*Proof: case 1.* $F$ has finite variance. Now results on order statistics imply that

$$Q_i - \bar{Q}_i \to 0 \text{ in probability, } i = 1, 2. \tag{7.11}$$

*Proof: case 2.* $F$ has finite variance. Now both $Q_2$ and $\bar{Q}_2$ tend to classify $F$ as heavy tailed with asymptotic probability unity, and hence suggest the same statistic, i.e. the Wilcoxon statistic.

Together with equation (7.11) this completes the proof. These propositions show that our adaptive procedures are asymptotically distribution free.

As the foregoing result is asymptotic, it is useful to verify it empirically. In addition, the powers of the adaptive tests, and the average lengths as well as the coverage probabilities of the adaptive confidence intervals, are also worth studying empirically. Since the efficiency results in the two-sample case carry over to the $c$-sample case, it is sufficient to consider the two-sample case. Accordingly we performed the following Monte Carlo studies.

Samples $X_1 = (X_{1,1}, \ldots, X_{1,20})$ and $X_2 = (X_{2,1}, \ldots, X_{2,20})$ were drawn 4000 times from each of the 12 distributions described later. The nominal level was set at $\alpha = 0.05$. The 2.5 % and 97.5 % quantiles of each test statistic were calculated using asymptotic normality. We shall say a test based on a statistic $h$ rejects $H_0$ for a particular sample if the value of $h$ based on that sample either falls below the 2.5 % quantile or exceeds the 97.5 % quantile.

For any underlying $F$, let $\hat{p}$ be the proportion of samples (in the 4000 samples) which lead to rejection of $H_0$. Suppose $\bar{\alpha}$ is 0.05. Then the normal approximation to the binomial shows that, with probability 95 %, $\hat{p}$ lies in the interval (0.043, 0.057). Conversely, therefore, if $\hat{p}$ lies in that interval then it is consistent with the hypothesis that $\bar{a} = 0.05$.

The empirical powers of these tests, under the alternative

$$\Delta = \frac{1}{\sqrt{(20+20)}} = \frac{1}{\sqrt{40}} = 0.158,$$

were studied next. More precisely, 0.158 was added to each member of the $x_2$ sample, and the proportion of times each test (based on $x_1$, $x_2 + 0.158$) rejected $H_0$ was computed. These proportions are the empirical powers and are also given in Table 4.

We then computed the average lengths and coverage probabilities of the confidence intervals, based on these tests. However, now we had to confine ourselves to the first

TABLE 4
*Empirical levels and powers of tests*

| *Distribution* | *Wilcoxon* | | *Adaptive I* | | *Adaptive II* | | *t test* | |
|---|---|---|---|---|---|---|---|---|
| | *Level* | *Power (%)* | *Level* | *Power (%)* | *Level* | *Power (%)* | *Level* | *Power (%)* |
| 1 | 5.23 | 49.48 | 5.45 | 63.3 | 5.3 | 70.9 | 5.2 | 50.1 |
| 2 | 5.23 | 38.1 | 5.5 | 44.23 | 5.55 | 46.23 | 5.2 | 41.48 |
| 3 | 5.05 | 12.43 | 5.13 | 15.98 | 5.2 | 18.53 | 4.8 | 12.23 |
| 4 | 5.23 | 12.2 | 5.4 | 11.95 | 5.45 | 11.9 | 5.2 | 12.65 |
| 5 | 5.25 | 11.5 | 5.4 | 11.4 | 5.42 | 11.45 | 3.38 | 7.2 |
| 6 | 5.27 | 12.3 | 5.35 | 11.5 | 5.35 | 11.4 | 5.15 | 10.38 |
| 7 | 5.23 | 8.7 | 5.15 | 8.5 | 5.15 | 8.5 | 3.45 | 4.8 |
| 8 | 4.45 | 19.13 | 4.45 | 25.18 | 4.38 | 25.2 | 4.45 | 12.03 |
| 9 | 4.95 | 14.23 | 5.05 | 16.88 | 5.05 | 17.48 | 4.53 | 12.13 |
| 10 | 5.2 | 27.18 | 5.3 | 42.08 | 5.3 | 42.08 | 5.33 | 11.53 |
| 11 | 5.15 | 14.95 | 4.9 | 19.5 | 4.9 | 19.75 | 5.03 | 8.98 |
| 12 | 5.1 | 32.2 | 5.08 | 39.88 | 5.08 | 39.93 | 4.2 | 9.8 |

TABLE 5
*Coverage probabilities and average lengths of confidence intervals*

| Distribution | Wilcoxon | | Adaptive I | | Adaptive II | | t test | |
|---|---|---|---|---|---|---|---|---|
| | Probability (%) | Length | Probability (%) | Length | Probability (%) | Length | Probability (%) | Length |
| 1 | 94.77 | 0.4 | 94.55 | 0.3 | 94.7 | 0.27 | 94.8 | 0.39 |
| 2 | 94.77 | 0.47 | 94.45 | 0.42 | 94.45 | 0.41 | 94.8 | 0.43 |
| 3 | 94.95 | 1.38 | 94.87 | 1.1 | 94.8 | 1.0 | 95.2 | 1.28 |
| 4 | 94.77 | 1.32 | 94.6 | 1.34 | 94.55 | 1.35 | 94.8 | 1.28 |
| 5 | 94.75 | 1.43 | 94.6 | 1.44 | 94.58 | 1.44 | 96.62 | 2.88 |
| 6 | 94.73 | 1.56 | 94.65 | 1.62 | 94.65 | 1.63 | 94.85 | 1.89 |
| 7 | 94.77 | 2.75 | 94.85 | 2.76 | 94.85 | 2.76 | 96.55 | 30.74 |
| 8 | 95.55 | 0.92 | 95.55 | 0.69 | 95.62 | 0.70 | 95.55 | 1.50 |
| 9 | 95.05 | 1.1 | 94.95 | 0.95 | 94.95 | 0.96 | 95.47 | 1.25 |
| 10 | 94.8 | 0.95 | 94.7 | 0.58 | 94.7 | 0.58 | 94.67 | 1.72 |
| 11 | 94.85 | 1.27 | 95.1 | 0.95 | 95.1 | 0.95 | 94.97 | 2.57 |
| 12 | 94.9 | 0.6 | 94.92 | 0.54 | 94.92 | 0.54 | 95.8 | 2.37 |

1000 samples ($X_1$, $X_2$), owing to restrictions on computer time. The results are displayed in Table 5.

Finally, we introduced the notion of efficiency of one procedure relative to another procedure, as the ratio of the reciprocals of the squared lengths of the corresponding confidence intervals. Then, the efficiency of $A$ relative to $B$ is $L_B^2/L_A^2$. The efficiencies of the adaptive procedures relative to the Wilcoxon procedure and Student's $t$ procedure are set out in Table 6.

We shall explain the notation for the distributions studied. $R$ denotes the uniform distribution with mean zero and variance unity. $N$ denotes the standard normal, while $0.95N + 0.05\ (10N)$ denotes the contaminated normal, with 5% contamination by a normal distribution with mean zero and variance 100. The log-normal distribution denotes the distribution of $\exp X$, where $X$ is standard normal. $G(2)$ denotes the gamma distribution with density $x \exp(-x)$ for $x \geqslant 0$, and zero, for $x < 0$. $W$ denotes the Weibull distribution with density $(1/2\sqrt{x}) \exp(-\sqrt{x})$ for $x \geqslant 0$ and zero for $x < 0$.

TABLE 6
*Efficiencies relative to the Wilcoxon and t tests*

| Distribution | Relative to the Wilcoxon test | | Relative to the t test | |
|---|---|---|---|---|
| | Adaptive I | Adaptive II | Adaptive I | Adaptive II |
| 1 | 1.77 | 2.19 | 1.69 | 2.09 |
| 2 | 1.25 | 1.31 | 1.05 | 1.1 |
| 3 | 1.57 | 1.90 | 1.35 | 1.64 |
| 4 | 0.97 | 0.96 | 0.91 | 0.90 |
| 5 | 0.99 | 0.99 | 4.0 | 4.0 |
| 6 | 0.93 | 0.92 | 1.36 | 1.34 |
| 7 | 0.99 | 0.99 | 120.84 | 120.84 |
| 8 | 1.78 | 1.73 | 4.73 | 4.59 |
| 9 | 1.34 | 1.31 | 1.73 | 1.70 |
| 10 | 2.68 | 2.68 | 8.79 | 9.79 |
| 11 | 1.79 | 1.79 | 7.32 | 7.32 |
| 12 | 1.23 | 1.23 | 19.27 | 19.27 |

The 12 distributions were $R$, the angular distribution having distribution function

$$
\begin{aligned}
F(x) &= 0, \quad x < -\frac{\pi}{4} \\
&= \left[\sin\left(\frac{\pi}{4}+x\right)\right]^2, \quad -\frac{\pi}{4} \leqslant x \leqslant \frac{\pi}{4} \\
&= 1, \quad x > \frac{\pi}{4},
\end{aligned}
$$

$0.9R + 0.1N$, $N$, $0.95N + 0.05\ (10N)$, double exponential, Cauchy, the standard exponential, $G(2)$, the chi-square distribution with one degree of freedom, log-normal and $W$.

The second distribution is useful in biomedical modelling (see Miller and Halpern (1980), p. 109), while the fifth distribution, the contaminated normal, arises frequently in practice (see Lehmann (1975)). The standard exponential distribution is used both in reliability and in biomedical sciences. The log-normal distribution arises in the atmospheric and biomedical sciences. The $W$ distribution is used in reliability theory.

Let us now analyse the findings of the Monte Carlo studies.

Table 4 shows that, for the adaptive tests, the empirical levels lie between the acceptable limits of 0.043 and 0.057. Therefore, for these tests the actual level can be assumed to be the same as the nominal level.

Next we consider the empirical powers. The adaptive procedures are slightly inferior to the $t$ procedure for the fourth distribution, and to the Wilcoxon procedure for the fourth to seventh distributions, but this is more than made up for by their superiority in the remaining cases.

Finally, we consider the efficiencies of the adaptive procedures relative to the Wilcoxon and $t$ procedures (see Table 6). Recall that this efficiency is defined in terms of the reciprocals of the squares of the average lengths of the confidence intervals. This definition makes sense, since all the confidence intervals have nearly the same coverage probability, i.e. almost always close to 95% (see Table 5).

These efficiencies are at worse about 0.9, but often far exceed unity. Thus, by employing these adaptive procedures instead of the traditional procedures (parametric or nonparametric) we may occasionally lose a little but often gain a lot.

## 8. Conclusion

This paper proposes two adaptive nonparametric procedures for multiple comparisons and testing for ordered alternatives in the one-way ANOVA model. Compared with the traditional normal theory procedure, and the nonparametric procedure based on ranks, these procedures are occasionally slightly inferior, but often considerably superior, especially for light-tailed distributions and skewed distributions. Since such distributions arise frequently in practice, these adaptive procedures can be safely recommended to applied statisticians.

## Acknowledgements

We are deeply grateful to the editor and the referees for valuable suggestions and comments and to Professor Prentice for granting us permission to use the lung cancer data sets in Section 6.

Madan Puri's research was supported by the Office of Naval Research, contract NO0014-85-K-0648.

## References


Bauer, D. F. (1972) Constructing confidence sets using rank statistics. *J. Amer. Statist. Ass.*, **67**, 687–690.

Chen, H. J. (1982) A new range statistic for comparisons of several exponential location parameters. *Biometrika*, **69**, 257–260.

Conover, W. J. (1973) Rank tests for one sample, two samples and $k$ samples without the assumption of a continuous distribution function. *Ann. Statist.*, **1**, 1105–1125.

Hajek, J. (1969) *A Course in Nonparametric Statistics.* San Francisco: Holden–Day.

Hajek, J. and Sidak, J. (1967) *Theory of Rank Tests.* New York: Academic Press.

Harter, L. H. (1960) *Order Statistics and Their Use in Testing and Estimation*, vol. I. Dayton: Aerospace Research Laboratories, US Air Force.

———(1974). Comment on Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *J. Amer. Statist. Ass.*, **69**, 923.

Hogg, R. V. (1974) Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *J. Amer. Statist. Ass.*, **69**, 909–926.

Hogg, R. V., Fisher, D. M. and Randles, R. H. (1975) A two-sample adaptive distribution-free test. *J. Amer. Statist. Ass.*, **70**, 656–661.

Lehmann, E. L. (1975) *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden–Day.

Miller, R. G., Jr, and Halpern, J. (1980) Robust estimators for quantal bioassay. *Biometrika*, **67**, 103–110.

Padmanabhan, A. R. (1977) Hodges–Lehmann estimators in the case of grouped data—I. *Commun. Statist.*, **6**, 371–380.

Puri, M. L. (1964) Asymptotic efficiency of a class of $c$-sample tests. *Ann. Math. Statist.*, **35**, 102–121.

———(1965) Some distribution-free $k$-sample rank tests of homogeneity against ordered alternatives. *Commun. Pure Appl. Math.*, **18**, 51–63.

Puri, M. L. and Sen, P. K. (1971) *Nonparametric Methods in Multivariate Analysis.* New York: Wiley.

Randles, R. H. and Hogg, R. V. (1973) Adaptive distribution-free tests. *Commun. Statist.*, 337–356.

Randles, R. H. and Wolfe D. A. (1979) *Introduction to the Theory of Nonparametric Statistics.* New York: Wiley.

Vorlickova, D. (1970) Asymptotic properties of rank tests under discrete distributions. *Z. Wahrch. Verw. Geb.*, **14**, 275–289.

Wegman, E. J. and Carroll, R. J. (1977) A Monte Carlo study of some robust estimators of location. *Commun. Statist.*, **6**, 795–812.