

Adaptive Nonparametric Tests for the Generalized Behrens-Fisher Problem

Uttam Bandyopadhyay¹, Atanu Biswas², and Dhiman Dutta¹

¹Department of Statistics, University of Calcutta, India

²Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

Abstract: Some adaptive test procedures are developed for the generalized Behrens-Fisher problem. The one having a deterministic approach is based on calculating a measure of symmetry from each sample and using them as a basis for choosing between the modified Wilcoxon-Mann-Whitney test (Fligner and Policello, 1981) and the modified Mood's median test (Fligner and Rust, 1982). The other one is a probabilistic approach which also uses a combination of the modified Wilcoxon-Mann-Whitney test and the modified Mood's median test according to an evidence of asymmetry provided by the p-value from the triples test for symmetry given in Randles, Fligner, Policello, and Wolfe (1980). This probabilistic approach is further modified by using a suitable function of the p-value from the triples test. A simulation study reveals that the modified procedure performs reasonably well in terms of power and attainment of the nominal size.

Zusammenfassung: Adaptive Testprozeduren werden für das generalisierte Behrens-Fisher Problem entwickelt. Eine mit einem deterministischen Ansatz beruht auf der Berechnung von Symmetriemaßen jeder Stichprobe. Beruhend auf diese Maße wählt man nun zwischen dem modifizierten Wilcoxon-Mann-Whitney Test (Fligner and Policello, 1981) und dem modifizierten Mood's Median Test (Fligner and Rust, 1982). Die andere beruht auf einem probabilistischen Ansatz der auch eine Kombination des modifizierten Wilcoxon-Mann-Whitney Test und des modifizierten Mood's Median Test verwendet, in Beachtung von Hinweisen der Asymmetrie durch den p-Wert des Triples Test für Symmetrie aus Randles et al. (1980). Dieser probabilistische Ansatz wird weiter modifiziert, indem eine passende Funktion des p-Wertes des Triples Test zum Einsatz kommt. Eine Simulationsstudie zeigt, dass die modifizierte Prozedur ziemlich gut funktioniert bezüglich der Macht und dem Einhalten des nominalen Niveaus.

Keywords: Adaptive Test Procedures, Modified Wilcoxon-Mann-Whitney Test, Modified Mood's Median Test, Triples Test, Deterministic Approach, Probabilistic Approach, p-Value.

1 Introduction

Consider the problem of testing the null hypothesis that the medians of two populations having continuous cumulative distribution functions (cdf's) are equal against one- or two-sided alternative. If it is assumed that the populations have the same shape with common scale, the usual nonparametric procedures, such as the Wilcoxon-Mann-Whitney

test (Mann and Whitney, 1947; Wilcoxon, 1945) and the Mood's median test (Mood, 1954), are exactly distribution-free under the null hypothesis of equal medians. The distribution-free property allows the formulation of an exact size α test. But, in some situations, the two populations may have different shapes even under the null hypothesis of equal medians. In such a case, it is natural to treat the problem as a generalization of the Behrens-Fisher problem. For the standard nonparametric tests the level will not be preserved for small or large sample sizes when populations have different shapes or variances, and hence nonparametric estimation of the unknown parameters involved in the null distribution of the test statistic is required to obtain a distribution-free test procedure.

The Wilcoxon-Mann-Whitney statistic, under the assumption of symmetry for the underlying distributions, is modified by Potthoff (1963) and Zaremba (1962) to obtain conservative tests, while asymptotically distribution-free tests are proposed by Zaremba (1962) and Fligner and Policello (1981) for testing the equality of two medians without making any assumptions on the shapes of the underlying populations. If both the underlying distributions are not symmetric, the various modifications of the Wilcoxon-Mann-Whitney test are no longer asymptotically distribution free and therefore may not maintain the nominal size. However, with violation of symmetry, conservative procedures are suggested by Hettmansperger (1973) and Hettmansperger and Malin (1975). An asymptotically distribution-free test based on Mood's median test is proposed by Hettmansperger and Malin (1975) by first constructing a conservative version of Mood's test and then using an estimate of the null variance of the test statistic to raise the significance level closer to the nominal level. A modification of the Kruskal-Wallis test (Kruskal, 1952; Kruskal and Wallis, 1952) is proposed by Fligner and Rust (1984), which is exactly distribution-free under the assumption of equally shaped populations and is asymptotically distribution-free when the populations are assumed to be symmetric with equal medians. Fligner and Rust (1982) put forward a modification of the Mood's median test which is exactly distribution-free when the populations have the same shape and is asymptotically distribution-free when they do not.

A robust solution to the generalized Behrens-Fisher problem is proposed by Babu and Padmanabhan (2002), which consists of bootstrapping an appropriately centered version of the Mann-Whitney statistic. Brunner and Munzel (2000) proposed a rank test where the asymptotic variance is estimated consistently by using the ranks of overall observations as well as the ranks within each sample. It is not assumed that the underlying cdf's are continuous so that data with arbitrary ties can be handled. A permutation test based on the studentized rank statistic of Brunner and Munzel (2000) is proposed by Neubert and Brunner (2007). A likelihood ratio test for this problem is suggested by Troendle (2002), where the number of parameters in the score equations is effectively reduced to one by using a recursive formula for the remaining parameters. Bandyopadhyay and Das (2004, 2005), for 0-1 scores, proposed some partially sequential test procedures using placement statistics based on a progressively censored scheme.

The purpose of the present paper is to provide adaptive test procedures for the generalized Behrens-Fisher problem and to compare them with the existing non-adaptive competitors. The first adaptive procedure (AD1) has a deterministic approach based on the idea of Hogg, Fisher, and Randles (1975). A simple measure of skewness is calculated for each sample to assess whether both the distributions are symmetric. If both the un-

derlying distributions are found to be symmetric the modified Wilcoxon-Mann-Whitney test is used, otherwise the modified Mood's median test is used. The second adaptive procedure (AD2) has a probabilistic approach, which uses the p-value from the triples test for symmetry given in Randles et al. (1980). A modification of the proposed probabilistic approach is also suggested using a suitable function of the p-value from the triples test. This proposed modified probabilistic approach (AD3) has relatively high power for nearly all cases, maintaining the level of significance reasonably well.

The content of the article is arranged in the following way. Section 2 introduces our adaptive test procedures. Section 3 provides an example illustrating the proposed procedures. In Section 4 some numerical computations are presented to get an idea about the relative performance of the proposed test procedures over the various competitors. Some relevant asymptotic properties are discussed in Section 5. Finally Section 6 gives a brief concluding remark.

2 Statement of the Problem and the Proposed Adaptive Tests

Let X_1, X_2, \dots, X_{n_1} and Y_1, Y_2, \dots, Y_{n_2} be independent random samples corresponding to the populations with continuous cdf's $F(x)$ and $G(y)$, respectively. Let θ_X and θ_Y denote, respectively, the unique medians of the X and Y populations. The problem considered here is to test

$$H_0 : \theta_X = \theta_Y$$

against a suitable composite alternative. For simplicity, we consider the one-sided alternative

$$H_1 : \theta_X < \theta_Y.$$

Such a testing problem, violating the assumption on equality of shapes or scales for two or more unknown continuous cdf's, can be viewed as a generalization of the very famous Behrens-Fisher problem. Note that the more restrictive assumption that $G(y) = F((y - \theta_Y)/\tau)$ has not been made. The assumption is unnecessary and would not simplify the procedures.

Before introducing the proposed adaptive tests we present a brief description of the modified Wilcoxon-Mann-Whitney statistic, suggested by Fligner and Policello (1981) for the generalized Behrens-Fisher problem. The modified procedures are still exactly distribution-free when the populations are identical and asymptotically distribution-free, under some mild conditions, when the populations have equal medians but different shapes. The modified Wilcoxon-Mann-Whitney statistic can be used to test the null hypothesis H_0 provided that the underlying populations are symmetric. Let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n_1)}$ and $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n_2)}$ denote the ordered X and Y observations, respectively. Here F and G are assumed to be symmetric. Let Q_i denote the rank of $X_{(i)}$ in the combined sample and define the placement of X_i in Y_j 's as $P_i = Q_i - i$, the number of Y 's less than $X_{(i)}$, $i = 1, 2, \dots, n_1$. P_i is called the placement of $X_{(i)}$ among Y_1, Y_2, \dots, Y_{n_2} . If $F_{n_1}(x)$ and $G_{n_2}(y)$ are the empirical cdf's of the X and Y samples, respectively, we note that $P_i = n_2 G_{n_2}(X_{(i)})$. Similarly, we define the placement of Y_j

by $S_j = n_1 F_{n_1}(Y_{(j)})$ for $j = 1, 2, \dots, n_2$. The Wilcoxon-Mann-Whitney statistic U is defined as the number of pairs (X_i, Y_j) with $X_i < Y_j$, $i = 1, 2, \dots, n_1$, $j = 1, 2, \dots, n_2$. A consistent estimate of $\text{var}(U)$ can be obtained as

$$\widehat{\text{var}}(U) = \sum_{j=1}^{n_2} (S_j - \bar{S})^2 + \sum_{i=1}^{n_1} (P_i - \bar{P})^2 + \bar{P}\bar{S},$$

where $\bar{S} = \frac{1}{n_2} \sum_{j=1}^{n_2} S_j$ and $\bar{P} = \frac{1}{n_1} \sum_{i=1}^{n_1} P_i$. Note that, since the placements are functions of the ranks, the standardized statistic

$$\hat{U} = \frac{U - n_1 n_2 / 2}{\sqrt{\widehat{\text{var}}(U)}}$$

is a rank statistic. Under H_0 , \hat{U} is asymptotically standard normal when $n_1, n_2 \rightarrow \infty$ in such a way that $n_1/N \rightarrow \lambda$, $N = n_1 + n_2$, $0 < \lambda < 1$.

We next discuss in brief the modification of Mood's median test for the generalized Behrens-Fisher problem proposed by Fligner and Rust (1982). The modified Mood test is asymptotically distribution-free over the broad null hypothesis H_0 considered here without surrendering any power to the original test under the usual nonparametric assumption. Let $Z_{(1)}, Z_{(2)}, \dots, Z_{(N)}$ denote the combined sample order statistics. Then M , the combined sample median, is defined to be $Z_{((N+1)/2)}$ for N odd and $\frac{1}{2}[Z_{(N/2)} + Z_{(N/2+1)}]$ for N even. Hence, writing

$$\varphi(a, b) = \begin{cases} 1 & \text{if } a > b, \\ \frac{1}{2} & \text{if } a = b, \\ 0 & \text{if } a < b, \end{cases}$$

we get Mood's median test statistic as

$$T = \sum \varphi(Y_i, M) / n_2.$$

Under H_0 , assuming that $F'(x) = f(x)$ and $G'(x) = g(x)$ exist, $\sqrt{n_2}(T - \frac{1}{2})$ is asymptotically normal with mean 0 and variance σ^2 when $n_1, n_2 \rightarrow \infty$, where

$$\sigma^2 = \frac{1}{4} \frac{\lambda(1 - \lambda + \lambda R^2)}{(1 - \lambda + \lambda R)^2}, \quad R = f(\theta)/g(\theta)$$

with λ defined earlier and $\theta = \theta_X = \theta_Y$. Setting $b_N = [\frac{1}{2}(N+1) + K_N]$, where $[\cdot]$ denotes the greatest integer function, and the sequence $\{K_N\}$ such that $K_N = o(N^{1/2} \log N)$ and $K_N(N^{1/4} \log N)^{-1} \rightarrow +\infty$ as $N \rightarrow \infty$, let us consider

$$L_N = Z_{(N-b_N)} \quad \text{and} \quad U_N = Z_{(b_N)}.$$

The modified Mood's median test, due to Fligner and Rust (1982), is then given by the variable $\hat{T} = \sqrt{n_2}(T - \frac{1}{2})/\hat{\sigma}$, where $\hat{\sigma}^2$ is a consistent estimate of σ^2 and is obtained by

$$\hat{\sigma}^2 = \begin{cases} \frac{n_1(n_2 + n_1 \hat{R}^2)}{4(n_2 + n_1 \hat{R})^2} & \text{if } G_{n_2}(U_N) - G_{n_2}(L_N) > 0, \\ \frac{1}{4} & \text{if } G_{n_2}(U_N) - G_{n_2}(L_N) = 0 \end{cases}$$

with $\hat{R} = \{F_{n_1}(U_N) - F_{n_1}(L_N)\} / \{G_{n_2}(U_N) - G_{n_2}(L_N)\}$. We now proceed to introduce the proposed adaptive test procedures. In this paper we consider $K_N = N^{1/2}$.

2.1 The Deterministic Approach

Here we propose an asymptotically distribution-free adaptive two-sample test for the generalized Behrens-Fisher problem having a deterministic approach. Hogg et al. (1975) developed their adaptive procedures only in the context of hypothesis testing in the usual nonparametric two-sample and one-sample problems. But their nice properties also extend to more general situations. Hogg's procedure, or some modification thereof, can be effectively employed in our present situation as well. Here a fairly easy classification scheme is used which merely attempts to assess whether both the underlying distributions are symmetric. We use the data itself to make such classification, and on the basis of that information select an appropriate test statistic for testing H_0 . For this purpose we use the following measure of symmetry

$$Q(x) = \frac{\bar{U}_\gamma(x) - \bar{M}_{0.5}(x)}{\bar{M}_{0.5}(x) - \bar{L}_\gamma(x)},$$

$$Q(y) = \frac{\bar{U}_\gamma(y) - \bar{M}_{0.5}(y)}{\bar{M}_{0.5}(y) - \bar{L}_\gamma(y)},$$

where $\bar{U}_\gamma(x)$, $\bar{M}_\gamma(x)$, $\bar{L}_\gamma(x)$ denote, respectively, the γn_1 largest middle and smallest combined order statistics corresponding to the X sample, and $\bar{U}_\gamma(y)$, $\bar{M}_\gamma(y)$, $\bar{L}_\gamma(y)$ denote that of the Y sample (Randles and Wolfe, 1979, p. 389).

This statistic not only has strong intuitive appeal as a measure of skewness, but there is also a good theoretical reason for considering it. Since $Q(x)$ and $Q(y)$ are the ratio of two linear functions of the order statistics, its convergence properties are better than some of the other measures of symmetry. Under the null hypothesis, the order statistics are the complete sufficient statistics for the common, but unknown cdf. Hence, by Basu's theorem on ancillary statistic, they are independent of every statistic which is distribution-free. In the present situation, both \hat{T} and \hat{U} are asymptotically distribution-free and hence not generally independent of the selector statistics. It can be shown, however, that the difference between the nominal level α and the actual level $\hat{\alpha}$ converges to zero as the sample size increases. Thus the adaptive procedure can be argued to be asymptotically distribution-free. Moreover such measures are location-free.

The proposed deterministic approach (AD1) will accordingly use the following classification scheme. When the data indicate that both the populations are symmetric, i.e. $Q(x), Q(y) \in J$, use the modified Wilcoxon-Mann-Whitney statistic \hat{U} , where J is an interval suitably chosen so that the overall adaptive procedure achieves good power maintaining the nominal level satisfactorily. When the data indicate that both the populations are not symmetric, i.e. at least one of $Q(x)$ and $Q(y)$ does not belong to the interval J , use the modified Mood's median test statistic \hat{T} . Hogg et al. (1975) considered the same measure of skewness as the selector statistic for their adaptive procedure, along with a measure of tailweight, by taking $\gamma = 0.05$. It should be noted that the value of the measure is 1 when the underlying distribution is symmetric. According to Hogg if the value of this statistic is greater than 2 then the right tail of the distribution seems to be longer than the left; i.e., there is an indication that the distribution is skewed to the right. If the value is less than 0.5, the sample indicates that the distribution may be skewed to the left.

In our simulation study we have considered small and moderate sample sizes like

$n_1 = n_2 = 20$ and $n_1 = 25, n_2 = 15$. For $n_2 = 15$ if we take $\gamma = 0.05$ then $\gamma n_2 = 0.75$ and consequently $\bar{U}_{0.05}$ and $\bar{L}_{0.05}$ do not seem to be meaningful enough. So in our case we take $\gamma = 0.10$. Different choices of the interval $J = (c, d)$ are examined with c and d around 0.5 and 2, respectively. The choice of the interval should be such that the power of the proposed adaptive procedure is significantly close to that of the best test for the distribution considered while maintaining the nominal significance level. From the simulation studies $J = (0.5, 2.3)$ is found to be the best choice in terms of the robustness and the power of the adaptive procedure. We can say that the assumption of symmetry is tenable if $Q(x)$ and $Q(y)$ lie between 0.5 and 2.3.

2.2 The Probabilistic Approach

In the present section we formulate an adaptive nonparametric procedure for the generalized Behrens-Fisher problem having a probabilistic or stochastic approach. Instead of prescribing a fixed level of significance we use p-value based randomized classification rule for the selection of an appropriate test statistic. A disadvantage with the deterministic adaptive procedure, based on Hogg's principle, is the discontinuous nature of the test selection method. The test selection is likely to be affected if the value of the selector statistic is near the boundary between two partitioning sets. Here a small change in the data may move the observed value of the selector statistic over the boundary. Moreover the boundaries are defined empirically. Although their partitioning may give good results, more objective and stochastic considerations may lead to a different, even better, adaptive procedures. The proposed probabilistic procedure consists in calculating some classification probabilities, based on the p-values of pretests, to decide on the appropriate test for the generalized Behrens-Fisher problem. These so called probabilistic or stochastic adaptive procedures are shown to be effective and yet computationally simple enough to appeal to the practicing statistician.

The proposed probabilistic approach is also a combination of the modified Wilcoxon-Mann-Whitney statistic and the modified Mood's median test statistic according to the evidence provided by the p-values of pretests for symmetry. For testing symmetry we make use of the well-known triples test presented by Randles et al. (1980). Before introducing the proposed probabilistic approach we present a brief review of the triples test. The null hypothesis for the triples test is that the underlying population is symmetric about θ against the alternative that it is asymmetric. Let

$$h(x_1, x_2, x_3) = \frac{1}{3}[\text{sgn}(x_1 + x_2 - 2x_3) + \text{sgn}(x_1 + x_3 - 2x_2) + \text{sgn}(x_2 + x_3 - 2x_1)],$$

where $\text{sgn}(x) = 1, 0, -1$ as $x >, =, < 0$. The triples test is then based on the U-statistic

$$\hat{\eta} = \frac{1}{\binom{n}{3}} \sum_{i < j < k} h(X_i, X_j, X_k). \quad (1)$$

Reject the null hypothesis of symmetry if $|V| > \tau_{\alpha/2}$, where $\tau_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution,

$$V = \frac{\sqrt{n}\hat{\eta}}{\hat{v}}$$

with \hat{v} being a consistent estimate of the variance of the U-statistic $\hat{\eta}$.

We now introduce the proposed adaptive rule (AD2) having a probabilistic approach. Let p_1 and p_2 denote, respectively, the p-values corresponding to observed $\hat{\eta}$ -values for the X and Y samples, viz., $\hat{\eta}_1$ and $\hat{\eta}_2$. These two p-values can be regarded as the amount of evidence against symmetry for each of the two underlying distributions. Thus we may consider $p = \min(p_1, p_2)$ as the amount of evidence against symmetry for both the underlying distributions. So whenever p_1 and p_2 are observed, perform a Bernoullian trial with probability of success $p = \min(p_1, p_2)$. If success occurs, use the \hat{U} test; otherwise, use the \hat{T} test. In other words our adaptive test rule is: Reject H_0 with probability p if $\hat{U} > \hat{u}_{\alpha, n_1, n_2}$ and with probability $(1 - p)$ if $\hat{T} > \hat{t}_{\alpha, n_1, n_2}$; or equivalently we may say: Accept H_0 with probability p if $\hat{U} \leq \hat{u}_{\alpha, n_1, n_2}$ and with probability $(1 - p)$ if $\hat{T} \leq \hat{t}_{\alpha, n_1, n_2}$, where $\hat{u}_{\alpha, n_1, n_2}$ and $\hat{t}_{\alpha, n_1, n_2}$ are the upper α -critical values for the \hat{U} and the \hat{T} tests, respectively.

2.3 The Modified Probabilistic Approach

The adaptive procedure considered here is a modification of the probabilistic approach. We know that p-value higher than 0.05 is considered to be statistically insignificant at 5% level of significance. Thus, for example, $p = 0.10$ will indicate that both p_1 and p_2 are greater than or equal to 0.10, i.e., both the distributions may be considered to be symmetric, and hence it is desirable to use the modified Wilcoxon-Mann-Whitney test with a higher probability. However, the proposed probabilistic approach uses the \hat{U} test with probability 0.10 and the \hat{T} test with probability 0.90. Here $p = \alpha = 0.05$ can be treated as complete dilemma. To overcome this drawback we consider a real-valued function $k(\cdot)$ satisfying the following conditions: i) k is monotone, non-decreasing, ii) $k(0) = 0$, iii) $k(0.05) = 0.5$ and iv) $k(1) = 1$. Then we propose our modified adaptive rule (AD3) as follows. Perform the random experiment with probability of success $p^* = \min(p_1^*, p_2^*)$, where $p_1^* = k(p_1)$, $p_2^* = k(p_2)$. The test rule will remain same as in the probabilistic approach with p replaced by p^* .

Various choices of k satisfying the above conditions may be obtained. For example, one may consider the cdf of a beta distribution with median at α , the desired level of significance. But we must be aware of the dangers associated with the use of such adaptive procedures. Even though the final test is performed at the desired level of significance α , in the overall testing procedure the actual level may be quite different from the nominal level of significance. So the robustness of the adaptive procedures should be carefully examined. For example, we may consider the triples test which is a highly recommended test for testing whether a continuous univariate population is symmetric. We have used it in the first stage of our probabilistic approach and the overall test have been found to be robust enough. But, when we use the same statistic as the selector statistic in the deterministic approach, the robustness property is not maintained. So the choice of the selector statistic for such an adaptive procedure should be made with extra care. Here, all the proposed adaptive procedures are only asymptotically distribution-free and so the actual level of the adaptive procedures may be slightly higher than the nominal level. Thus, for the modified probabilistic approach, one should be concerned with an appropriate choice of k so that the adaptive procedure is robust and performs reasonably well in terms of power. The use of beta distribution with indices (1, 13.513406) fits our present situation.

3 An Example

As an example to illustrate the proposed methods we consider the data given in McNabb (2004, p. 242). A political scientist was interested in knowing the percentage of voters who supported a bill to limit increases in property taxes to no more than 5% per year. Data were collected in 28 countries out of which 15 countries were considered to be predominantly urban while 13 countries were predominantly rural. The null hypothesis for this study was that rural voters are no more likely to support the tax-limit bill than the urban voters whereas the alternative hypothesis was that rural voters are more likely to support limits on property taxes than are urban voters. The results of the telephone survey are given in Table 1.

Table 1: Percentage of voters supporting a tax-limit bill.

Urban countries	22.20, 19.90, 42.09, 26.12, 41.11, 46.44, 63.67, 44.12, 44.22, 44.23, 60.56, 33.12, 51.07, 43.07, 43.55
Rural countries	33.30, 29.89, 59.76, 35.22, 51.98, 54.66, 69.09, 45.24, 47.93, 53.22, 61.12, 42.90, 58.43

For testing $H_0 : \theta_X = \theta_Y$ against the alternative $H_1 : \theta_X < \theta_Y$ using the \hat{U} test, we reject H_0 in favor of H_1 if and only if $\hat{U} > \hat{u}_{\alpha, n_1, n_2}$ or in other words the corresponding p-value is less than the desired level of significance α . The p-value corresponding to \hat{U} test is 0.0426 and therefore we reject the null hypothesis H_0 with this data using the \hat{U} test at 5% level of significance. Using the \hat{T} test with significance level $\alpha = 0.05$ we reject H_0 in favor of H_1 if and only if $\hat{T} > \hat{t}_{\alpha, n_1, n_2}$ or equivalently the corresponding p-value is less than 0.05. For this data the p-value corresponding to \hat{T} test comes out to be 0.0586. So on the basis of the given data we accept H_0 at 5% level of significance using the \hat{T} test.

Clearly there is difference in decision between the two tests. So we may now proceed to illustrate the application of the proposed adaptive procedures. We first consider the AD1 test. For this we calculate $Q(x) = 0.9275325$ and $Q(y) = 0.8336492$. Both the observed $Q(x)$ and $Q(y)$ lie in the interval (0.5, 2.3). Thus based on this measure we may assume that both the underlying distributions are symmetric. Thus we use the \hat{U} test and hence we reject H_0 at 5% level of significance.

To perform the AD2 test we need to compute the p-values of the triples test for each of the two samples. The observed p_1 and p_2 values are 0.8030368 and 0.5041223 respectively. We now perform a Bernoullian trial with probability of success $p = \min(p_1, p_2) = 0.5041223$. If success occurs use the \hat{U} test, otherwise use the \hat{T} test.

It can be observed that both the p-values are much higher than 0.05 so we may consider both the underlying distributions to be symmetric. Thus it is desirable to use the \hat{U} test with a very high probability. However in the previous approach we used the \hat{U} test with probability 0.5041223. To overcome this drawback we consider the AD3 test. It has already been pointed out that the cdf of the beta distribution with indices (1, 13.513406) is considered as the choice of the function $k(\cdot)$. So here $p_1^* = 1$, $p_2^* = 0.9999235$ and thus

$p^* = \min(p_1^*, p_2^*) = 0.9999235$. We now perform a Bernoullian trial with probability of success p^* . If success occurs use the \hat{U} test, otherwise use the \hat{T} test.

4 Relative Comparisons of the Competing Tests

In this section we present the results of a simulation study to assess the relative performance of the proposed adaptive procedures with the existing non-adaptive competitors. The distributions that are used to generate the data for this simulation study are members of the generalized lambda family of distributions discussed in Ramberg and Schmeiser (1974). This family provides a wide range of distributions that are easily generated since they are defined in terms of the inverse of the cdf

$$F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1 - u)^{\lambda_4}] / \lambda_2, \quad 0 < u < 1,$$

where λ_1 and λ_2 are the location and scale parameters, respectively. The parameters λ_3 and λ_4 determine the shape of the distribution. The generalized lambda distributions are used because the skewness ($\alpha_3 = E(X - \mu)^3 / \sigma^3$) and kurtosis ($\alpha_4 = E(X - \mu)^4 / \sigma^4$) could be specified. The distributions used in this study are so chosen so that their skewness and kurtosis would cover a wide range of possibilities for the underlying distribution. The parameters defining the 10 selected distributions along with the associated skewness and kurtosis values are given in Table 2. The first 5 distributions considered are symmetric while the remaining 5 distributions are skewed.

Table 2: Distributions used in the simulation study.

Distribution	λ_1	λ_2	λ_3	λ_4	α_3	α_4
1	0	2	1	1	0	1.8
2	0	0.197454	0.134915	0.134915	0	3.0
3	0	-1	-0.08	-0.08	0	6.0
4	0	-1	-0.24	-0.24	0	126.0
5	0	-5.11256	-1	-1	does not exist	does not exist
6	0	1	0.00007	0.1	1.5	5.8
7	0	-1	-0.0075	-0.03	1.5	7.5
8	0	-1	-0.1	-0.18	2.0	21.2
9	0	-1	-0.001	-0.13	3.2	23.8
10	0	-1	-0.001	-0.17	3.9	40.7

We consider the three proposed adaptive procedures AD1, AD2 and AD3, and the other tests included in the study are the two component tests of the adaptive rule, i.e., the \hat{T} test and the \hat{U} test. The whole computation is carried out taking the nominal level of significance to be $\alpha = 0.05$. The entries given in Tables 3-6 are generated via computer simulations with 5,000 replications for each configuration. The empirical size and the power of the tests are computed as the relative frequency with which a particular test rejects the null hypothesis H_0 . We investigate the powers at $\theta_Y = \xi_{0.5}, \xi_{0.6}, \xi_{0.7}$, where ξ_q is the q th quantile of the distribution of Y . The results presented in Tables 3 and 5 are

computed assuming equal scale for both the underlying distributions, i.e., the simulations are carried out using two generalized lambda distributions having equal scale and shape parameters. Whereas those in Tables 4 and 6 are computed without the assumption of equality of the scale parameters of the two underlying distributions. Here the simulations are performed using two generalized lambda distributions keeping λ_3, λ_4 fixed but varying λ_2 . In fact in our study the scale parameter of the second population is taken to be twice that of the first population.

Table 3: Empirical size and power of the tests for $n_1 = n_2 = 20$ (equal scale).

Distribution	q	\hat{T}	\hat{U}	AD1	AD2	AD3
1	0.5	0.052	0.053	0.050	0.056	0.056
	0.6	0.171	0.275	0.264	0.181	0.258
	0.7	0.377	0.647	0.633	0.424	0.591
2	0.5	0.054	0.049	0.051	0.056	0.050
	0.6	0.165	0.191	0.189	0.167	0.184
	0.7	0.385	0.467	0.461	0.407	0.446
3	0.5	0.051	0.049	0.053	0.053	0.050
	0.6	0.161	0.177	0.169	0.164	0.174
	0.7	0.387	0.425	0.421	0.395	0.410
4	0.5	0.052	0.051	0.049	0.054	0.047
	0.6	0.163	0.170	0.164	0.170	0.166
	0.7	0.384	0.410	0.387	0.385	0.407
5	0.5	0.053	0.054	0.051	0.049	0.049
	0.6	0.165	0.131	0.148	0.148	0.133
	0.7	0.398	0.298	0.363	0.388	0.319
6	0.5	0.051	0.047	0.049	0.055	0.057
	0.6	0.181	0.283	0.206	0.185	0.205
	0.7	0.455	0.709	0.516	0.461	0.519
7	0.5	0.053	0.049	0.054	0.054	0.055
	0.6	0.172	0.209	0.206	0.174	0.195
	0.7	0.443	0.575	0.542	0.462	0.513
8	0.5	0.051	0.052	0.050	0.054	0.046
	0.6	0.163	0.187	0.187	0.165	0.180
	0.7	0.426	0.485	0.478	0.427	0.467
9	0.5	0.050	0.053	0.052	0.053	0.054
	0.6	0.177	0.284	0.195	0.181	0.201
	0.7	0.464	0.722	0.497	0.467	0.509
10	0.5	0.052	0.053	0.054	0.058	0.054
	0.6	0.178	0.298	0.196	0.185	0.205
	0.7	0.454	0.728	0.494	0.472	0.503

From the results of the simulation study, we may conclude that all the three proposed adaptive test procedures AD1, AD2 and AD3 are robust for nearly all cases. The proposed modified probabilistic approach AD3 although tends to be slightly anti-conservative in a few situations but it is reasonably better than the other two proposed adaptive procedures

in terms of the total error. On the other hand, the \hat{U} test does not hold its nominal level very well when both the underlying distributions are not symmetric. Thus it is not justifiable to include it in any power comparisons as the high powers might easily be attributed to the inflated levels. However, all the proposed adaptive procedures perform better than the \hat{T} test in all the situations except for the distribution 5 (the lambda approximation to the Cauchy distribution), where the \hat{T} test is the best test. The proposed deterministic approach AD1 and the modified probabilistic approach AD3 have the nearest power to the \hat{U} test which is the best test when both the underlying distributions are symmetric except for the situation already mentioned. The proposed probabilistic approach AD2 achieves the desired level of significance more closely than the modified approach but as expected the modified AD3 test is much more powerful than the AD2 test. In a number of situations the proposed AD3 test is even more powerful than the AD1 test. Thus, overall we may say that there is not much difference between the AD1 test and the AD3 test in terms of power, and hence these two tests seems to be more preferable compared to the other existing competitors for the generalized Behrens-Fisher problem.

5 Some Asymptotics

We now consider some asymptotic properties of the proposed adaptive test statistics. As the asymptotics for all the three proposed procedures are same, we consider here the test based on AD2 only. We know that the modified Mood's median test statistic

$$\hat{T} = \frac{\sqrt{N}(T - 1/2)}{\hat{\sigma}}$$

is asymptotically standard normal. Again if the underlying distribution is symmetric then we know that the null distribution of the modified Wilcoxon-Mann-Whitney statistic

$$\hat{U} = \frac{U - n_1 n_2 / 2}{\sqrt{\widehat{\text{var}}(U)}}$$

is asymptotically standard normal. Also note that the magnitude of the p-values of the triples tests involved in the proposed test is an indicator of the amount of asymmetry for each of the underlying distribution present in the two samples. In order to study the asymptotic properties of the proposed adaptive procedures we need to look into the asymptotic behavior of the p-values of the preliminary tests. The p-value for the triples test is defined as $P_{H_0}(|V| > v)$, where v is the observed value of V . Since the triples test statistic V is asymptotically standard normal, we can approximate the p-value by $2(1 - \Phi(v))$, where $\Phi(v)$ is the standardized normal cdf

The p-value is itself a random variable whose distribution, for the null hypothesis, is asymptotically uniform over $(0, 1)$. Moreover, for the alternative hypothesis, the p-value goes to zero with probability one as the sample size becomes large. We know that, in general, convergence in distribution does not imply convergence of the corresponding expected value. However, for every uniformly bounded continuous function g , convergence in distribution of the sequence of random variables $\{W_n\}$ to the random variable W does

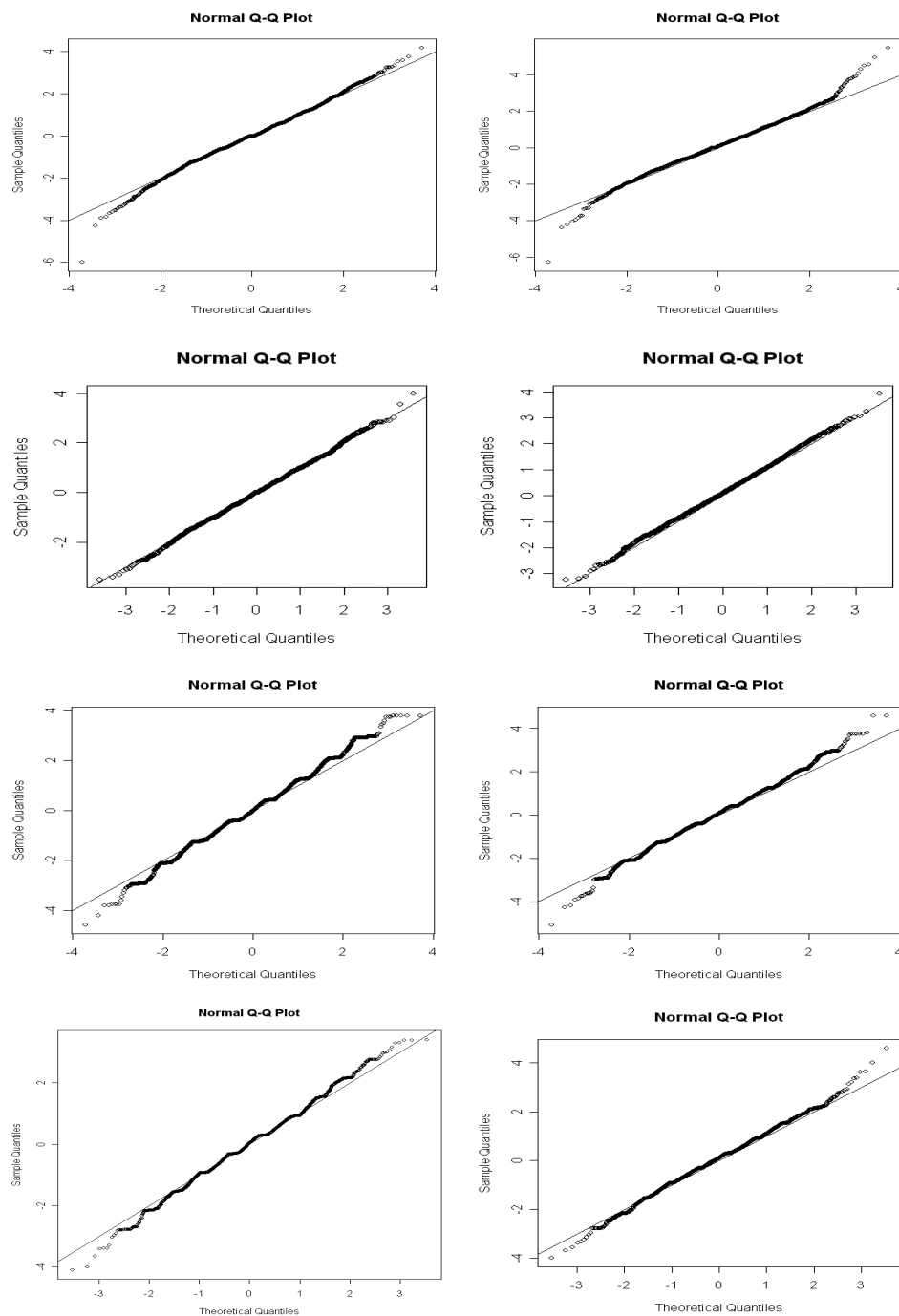


Figure 1: Normal QQ plots for distribution 3 (left) and distribution 8 (right) for $n_1 = n_2 = 20$ (first row), $n_1 = n_2 = 30$ (second row), $n_1 = 25$, $n_2 = 15$ (third row) and $n_1 = 35$, $n_2 = 25$ (fourth row).

Table 4: Empirical size and power of the tests for $n_1 = n_2 = 20$ (unequal scale).

Distribution	q	\hat{T}	\hat{U}	AD1	AD2	AD3
1	0.5	0.053	0.052	0.048	0.047	0.054
	0.6	0.186	0.273	0.259	0.204	0.254
	0.7	0.465	0.697	0.686	0.508	0.649
2	0.5	0.051	0.048	0.047	0.055	0.054
	0.6	0.202	0.246	0.238	0.208	0.229
	0.7	0.482	0.612	0.606	0.516	0.580
3	0.5	0.049	0.054	0.050	0.050	0.055
	0.6	0.204	0.231	0.220	0.205	0.206
	0.7	0.489	0.570	0.565	0.501	0.557
4	0.5	0.052	0.048	0.048	0.052	0.047
	0.6	0.195	0.223	0.208	0.197	0.212
	0.7	0.503	0.560	0.535	0.508	0.545
5	0.5	0.052	0.052	0.053	0.049	0.048
	0.6	0.203	0.168	0.191	0.185	0.173
	0.7	0.525	0.435	0.485	0.507	0.447
6	0.5	0.052	0.084	0.055	0.050	0.056
	0.6	0.179	0.350	0.210	0.183	0.209
	0.7	0.464	0.734	0.534	0.486	0.536
7	0.5	0.052	0.062	0.052	0.051	0.060
	0.6	0.195	0.286	0.249	0.199	0.238
	0.7	0.494	0.689	0.601	0.506	0.570
8	0.5	0.051	0.062	0.053	0.053	0.056
	0.6	0.201	0.253	0.248	0.204	0.235
	0.7	0.502	0.637	0.604	0.493	0.575
9	0.5	0.050	0.093	0.050	0.050	0.055
	0.6	0.180	0.362	0.196	0.185	0.209
	0.7	0.474	0.737	0.500	0.479	0.523
10	0.5	0.048	0.105	0.050	0.046	0.058
	0.6	0.183	0.368	0.194	0.186	0.200
	0.7	0.477	0.745	0.491	0.485	0.521

imply convergence of $E[g(W_n)]$ to $E[g(W)]$. Thus, as the sample size increases, the expected p-value here approaches 1 and 0 under the null and the alternative hypotheses, respectively. Hereinafter, whenever we discuss about the limiting p-values we indeed refer to the limit of the expected p-values. If the data do not provide sufficient evidence of asymmetry for at least one of the samples, then $p = \min(p_1, p_2)$ approaches 1, otherwise it tends to 0.

Keeping the above results in mind we now study the asymptotic behavior of

$$AD2 = \hat{U}I(U^* \leq p) + \hat{T}I(U^* > p),$$

where $I(x)$ is an indicator function assuming the values 1 or 0 according as x is true or false and U^* is uniformly distributed over $(0, 1)$ and is independent of $\{X_1, X_2, \dots, X_{n_1}, Y_1, Y_2, \dots, Y_{n_2}\}$.

Table 5: Empirical size and power of the tests for $n_1 = 25$, $n_2 = 15$ (equal scale).

Distribution	q	\hat{T}	\hat{U}	AD1	AD2	AD3
1	0.5	0.037	0.053	0.046	0.046	0.043
	0.6	0.115	0.259	0.219	0.159	0.246
	0.7	0.273	0.617	0.556	0.403	0.591
2	0.5	0.038	0.052	0.048	0.044	0.045
	0.6	0.111	0.185	0.172	0.171	0.184
	0.7	0.284	0.455	0.425	0.332	0.428
3	0.5	0.038	0.046	0.043	0.038	0.051
	0.6	0.111	0.169	0.154	0.133	0.161
	0.7	0.279	0.413	0.383	0.324	0.397
4	0.5	0.036	0.050	0.044	0.043	0.054
	0.6	0.109	0.152	0.136	0.128	0.143
	0.7	0.280	0.385	0.346	0.303	0.367
5	0.5	0.037	0.054	0.042	0.035	0.044
	0.6	0.102	0.127	0.108	0.105	0.107
	0.7	0.284	0.295	0.292	0.284	0.287
6	0.5	0.036	0.050	0.038	0.035	0.041
	0.6	0.143	0.282	0.157	0.145	0.172
	0.7	0.451	0.727	0.473	0.455	0.511
7	0.5	0.037	0.049	0.047	0.038	0.046
	0.6	0.135	0.207	0.167	0.135	0.175
	0.7	0.382	0.571	0.457	0.397	0.476
8	0.5	0.035	0.050	0.045	0.041	0.046
	0.6	0.120	0.176	0.154	0.142	0.164
	0.7	0.337	0.465	0.410	0.372	0.438
9	0.5	0.041	0.050	0.041	0.035	0.038
	0.6	0.150	0.294	0.160	0.150	0.161
	0.7	0.487	0.749	0.497	0.490	0.508
10	0.5	0.037	0.051	0.044	0.038	0.042
	0.6	0.153	0.295	0.163	0.156	0.160
	0.7	0.491	0.759	0.505	0.499	0.506

At the first stage a normal QQ (Quantile-Quantile) plot is used to check the asymptotic normality of AD2 under H_0 . The normal QQ plot is a very useful visual tool for assessing whether the distribution of a given variable follows a normal distribution. The QQ plot plots the empirical quantiles against the theoretical quantiles for normal distribution. When the distribution of the variable under examination has the same shape as the reference distribution, the normal distribution in this case, the QQ plot is linear. We have generated QQ plots using some of the 10 selected members of the generalized lambda family and present here the normal QQ plots for the distribution 3, which is symmetric and has a moderate tailweight, along with the distribution 8, which is positively skewed with heavier tailweight. All the normal QQ plots presented here are generated using two distributions with unequal scale parameters. The normal QQ plots seems to be fairly

Table 6: Empirical size and power of the tests for $n_1 = 25$, $n_2 = 15$ (unequal scale).

Distribution	q	\hat{T}	\hat{U}	AD1	AD2	AD3
1	0.5	0.050	0.048	0.047	0.051	0.047
	0.6	0.215	0.295	0.266	0.244	0.290
	0.7	0.510	0.735	0.696	0.607	0.726
2	0.5	0.050	0.052	0.052	0.055	0.050
	0.6	0.210	0.249	0.246	0.226	0.243
	0.7	0.525	0.644	0.629	0.568	0.626
3	0.5	0.051	0.048	0.055	0.053	0.049
	0.6	0.212	0.233	0.227	0.214	0.225
	0.7	0.523	0.606	0.581	0.548	0.585
4	0.5	0.049	0.048	0.047	0.048	0.048
	0.6	0.202	0.227	0.218	0.202	0.226
	0.7	0.522	0.568	0.549	0.540	0.562
5	0.5	0.048	0.046	0.053	0.047	0.045
	0.6	0.195	0.171	0.189	0.179	0.162
	0.7	0.511	0.439	0.497	0.491	0.452
6	0.5	0.052	0.080	0.050	0.054	0.058
	0.6	0.231	0.380	0.239	0.247	0.253
	0.7	0.618	0.780	0.637	0.629	0.663
7	0.5	0.053	0.069	0.054	0.050	0.055
	0.6	0.228	0.310	0.262	0.244	0.262
	0.7	0.598	0.727	0.638	0.617	0.664
8	0.5	0.051	0.055	0.051	0.054	0.060
	0.6	0.229	0.266	0.238	0.234	0.253
	0.7	0.579	0.664	0.617	0.598	0.645
9	0.5	0.052	0.091	0.057	0.053	0.053
	0.6	0.244	0.390	0.248	0.254	0.257
	0.7	0.649	0.803	0.654	0.650	0.659
10	0.5	0.054	0.099	0.054	0.056	0.058
	0.6	0.238	0.395	0.257	0.239	0.245
	0.7	0.648	0.797	0.668	0.650	0.659

linear even for the combined sample size $N = 40$ with $n_1 = n_2 = 20$ and $n_1 = 25$, $n_2 = 15$. It is also observed that as we increase the combined sample size to $N = 60$ with $n_1 = n_2 = 30$ and $n_1 = 35$, $n_2 = 25$ the normal QQ plots tends to be much more linear. Thus the normal QQ plots provide us with a fair indication about the asymptotic normality of AD2 under H_0 . We now proceed to verify the asymptotic normality of AD2 under H_0 theoretically in the following result:

Result 1. For $\min(n_1, n_2) \rightarrow \infty$, the statistic AD2 has asymptotically standard normal distribution under H_0 .

Proof. Let $\psi_N(\tau)$ denote the cdf corresponding to AD2 under H_0 . Further we denote the events $[\hat{U} \leq \tau]$, $[\hat{T} \leq \tau]$ and $[U^* \leq p]$ by A_N , B_N and E_N , respectively. Then we can

write

$$\psi_N(\tau) = P_{H_0}(AD2 \leq \tau) = P_{H_0}(A_N \cap E_N) + P_{H_0}(B_N \cap E_N^c).$$

We know that,

$$\lim_{n_1, n_2 \rightarrow \infty} P_{H_0}(E_N) = 1 \text{ or } 0 \quad (2)$$

according as each of the two underlying populations are symmetric or otherwise. If both the underlying distributions are symmetric then $\eta_1 = \eta_2 = 0$, where $\eta_i = E(\hat{\eta}_i)$, $i = 1, 2$. Also, note that

$$\lim_{n_1, n_2 \rightarrow \infty} P_{H_0}(A_N) = \Phi(\tau), \quad \text{provided } \eta_1 = \eta_2 = 0 \quad (3)$$

and

$$\lim_{n_1, n_2 \rightarrow \infty} P_{H_0}(B_N) = \Phi(\tau), \quad \text{whatever } \eta_1 \text{ and } \eta_2 \text{ may be.} \quad (4)$$

We now consider the following two cases:

Case 1. When $\eta_1 = \eta_2 = 0$

$$P_{H_0}(B_N \cap E_N^c) \leq P_{H_0}(E_N^c) \rightarrow 0 \quad \text{as } \min(n_1, n_2) \rightarrow \infty$$

Thus

$$\lim_{n_1, n_2 \rightarrow \infty} P_{H_0}(B_N \cap E_N^c) = 0.$$

Moreover, combining (2), (3) and

$$P_{H_0}(A_N) + P_{H_0}(E_N) - 1 \leq P_{H_0}(A_N \cap E_N) \leq \min(P_{H_0}(A_N), P_{H_0}(E_N)),$$

we find that

$$\lim_{n_1, n_2 \rightarrow \infty} P_{H_0}(A_N \cap E_N) = \Phi(\tau).$$

Hence

$$\lim_{n_1, n_2 \rightarrow \infty} \psi_N(\tau) = \Phi(\tau). \quad (5)$$

Case 2. When at least one $\eta_i \neq 0$, $i = 1, 2$

$$P_{H_0}(A_N \cap E_N) \leq P_{H_0}(E_N) \rightarrow 0 \quad \text{as } \min(n_1, n_2) \rightarrow \infty,$$

which gives

$$\lim_{n_1, n_2 \rightarrow \infty} P_{H_0}(A_N \cap E_N) = 0.$$

Again, combining (2), (4) and

$$P_{H_0}(B_N) + P_{H_0}(E_N^c) - 1 \leq P_{H_0}(B_N \cap E_N^c) \leq \min(P_{H_0}(B_N), P_{H_0}(E_N^c))$$

we obtain

$$\lim_{n_1, n_2 \rightarrow \infty} P_{H_0}(B_N \cap E_N^c) = \Phi(\tau).$$

Thus,

$$\lim_{n_1, n_2 \rightarrow \infty} \psi_N(\tau) = \Phi(\tau). \quad (6)$$

Finally, combining (5) and (6) we get

$$\lim_{n_1, n_2 \rightarrow \infty} \psi_N(\tau) = \Phi(\tau),$$

without any assumption regarding the symmetry of the two underlying distributions. \square

The asymptotic power properties of the adaptive test procedures depend on the criteria and the statistic used. We focus our attention on the case where F and G belong to the same location and scale family. Let X_1, X_2, \dots, X_{n_1} and Y_1, Y_2, \dots, Y_{n_2} be independent random samples corresponding to the populations with continuous cdf's $F(x) = H(x - \theta_X)$ and $G(y) = H\{(y - \theta_Y)/\tau\}$ respectively, where $H(\cdot)$ is an arbitrary continuous cdf with $H(0) = 1/2$. Suppose $H(x)$ has the density $h(x)$ at all real x . Then under the sequence of Pitman local alternatives

$$\theta_Y = \theta_X + b/\sqrt{N} \quad (7)$$

and suitable regularity conditions, the asymptotic power of the modified Mood's test is given by $\Phi(e_1 b - \tau_\alpha)$ and that of the modified Wilcoxon-Mann-Whitney test is given by $\Phi(e_2 b - \tau_\alpha)$, where

$$e_1 = 2 \frac{\{\lambda(1-\lambda)\}^{1/2} h(0)}{(1-\lambda) + \lambda\tau^2}, \quad e_2 = \{\lambda(1-\lambda)\}^{1/2} \frac{\int h(x/\tau)h(x)dx}{\tau\{\lambda\phi + (1-\lambda)\gamma\}^{1/2}},$$

$$\phi = \int F^2 dG - \left(\int F dG \right)^2, \quad \gamma = \int G^2 dF - \left(\int G dF \right)^2.$$

We see that power of the adaptive test converges to the power of the better component. Thus under (7) the power of each of the proposed adaptive test procedures converges to the power of the \hat{U} test when both the underlying distributions are symmetric, and to that of the \hat{T} test otherwise. This is shown in the following result.

Result 2. *The asymptotic power of the adaptive test AD2 under the sequence of local alternatives (7) is*

$$I(\eta_1^2 + \eta_2^2 = 0)\Phi(e_2 b - \tau_\alpha) + \{1 - I(\eta_1^2 + \eta_2^2 = 0)\}\Phi(e_1 b - \tau_\alpha).$$

Proof. Using the same technique as in Result 1, the asymptotic power of the upper α level AD2 test under (7) can be obtained as

$$\begin{aligned} \beta(b) &= \lim_{n_1, n_2 \rightarrow \infty} P(AD2 > \tau_\alpha) \\ &= I(\eta_1^2 + \eta_2^2 = 0)\Phi(e_2 b - \tau_\alpha) + \{1 - I(\eta_1^2 + \eta_2^2 = 0)\}\Phi(e_1 b - \tau_\alpha). \end{aligned}$$

\square

6 Concluding Remarks

In the nonparametric two-sample location problem, the most common statistical problem is the testing of the null hypothesis that the two populations have equal medians against one-or two-sided alternatives. Here the only assumption is that the two populations are continuous. In this paper we have developed some adaptive procedures for testing the so called generalized Behrens-Fisher problem. The deterministic approach (AD1) proposed in this paper is based on calculating a simple measure of symmetry for each sample involving the terms $\bar{U}_\gamma(x)$, $\bar{M}_\gamma(x)$, $\bar{L}_\gamma(x)$ and $\bar{U}_\gamma(y)$, $\bar{M}_\gamma(y)$, $\bar{L}_\gamma(y)$ which denote the averages of the γn_1 and γn_2 largest, middle and smallest combined order statistics for the X and Y samples, respectively. If each of these measures falls in a specific interval then both the underlying distributions are considered to be symmetric. A disadvantage with this deterministic approach is that both the choice of γ and that of the interval are made subjectively.

In the proposed probabilistic approach we make use of the triples test which is asymptotically distribution-free for testing whether the distribution is symmetric about some unknown location parameter. We want to compare between our proposed p-value based randomized classification rule and the deterministic rule based on the concept of Hogg. So in the proposed deterministic approach we use the same selector statistic as in Hogg et al. (1975) to assess the skewness of the underlying distribution.

The simulation study demonstrates that the proposed probabilistic approach AD2 is robust for nearly all the situations but the proposed deterministic approach AD1 and the modified probabilistic approach AD3, being reasonably robust, have relatively much higher powers. We have already pointed out the subjectivity in the proposed AD1 test and we recommend to use the AD3 test for the two-sample generalized Behrens-Fisher problem. The result seems to be important for statistical applications, because a practicing statistician has usually little or no information about the underlying distribution of the data. Adaptive tests are not designed to be optimal for any particular distribution but this study convinces us that they are certainly worth considering in practical problems. Note that, although to perform the simulation study and obtain the expression for the asymptotic power of the proposed adaptive test it is assumed that F and G belong to the same location and scale family they are also valid when F and G do not belong to the same location and scale family. It should also be noted that we can construct adaptive tests for the multisample generalized Behrens-Fisher problem.

In the modified probabilistic approach AD3 it is argued that the random experiment should be done with probability $p^* = \min(p_1^*, p_2^*)$, where $p_1^* = k(p_1)$, $p_2^* = k(p_2)$ and $k(\cdot)$ is such that i) k is monotone, non-decreasing, ii) $k(0) = 0$, iii) $k(0.05) = 0.5$ and iv) $k(1) = 1$. For example we can take $k(\cdot)$ as the cdf of a suitable beta distribution with median at α . The median of $\text{beta}(\beta_1, \beta_2)$ is taken as α . We fix β_1 , and find β_2 . For different choices of β_1 there will be different β_2 for the fixed α . Thus we obtain different possible choices of p^* for different choices of β_1 . Here we have used one such choice of β_1 and β_2 . However it may be possible to find an optimal β_1 using some criterion. We defer this to future research.

Acknowledgements:

The authors would like to thank the editor and the anonymous referee for their careful review of the manuscript. The third author is also grateful to the Council of Scientific and Industrial Research, India, for providing financial support under CSIR sanction no. 09/028(0694)/2007-EMR-I.

References

- Babu, G. J., and Padmanabhan, A. R. (2002). Resampling methods for the nonparametric Behrens-Fisher problem. *Sankhya, Series A*, 64, 678-692.
- Bandyopadhyay, U., and Das, R. (2004). A progressive censored test for generalized Behrens-Fisher problem. *Nonparametric Statistics*, 16, 731-752.
- Bandyopadhyay, U., and Das, R. (2005). Partial sequential test for generalized Behrens-Fisher problem. *Sequential Analysis*, 24, 365-387.
- Brunner, E., and Munzel, U. (2000). The nonparametric Behrens-Fisher problem: asymptotic theory and a small-sample approximation. *Biometrical Journal*, 42, 17-25.
- Fligner, M. A., and Policello, G. E. (1981). Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association*, 81.
- Fligner, M. A., and Rust, S. W. (1982). A modification of the Mood's median test for the generalized Behrens-Fisher problem. *Biometrika*, 69, 221-226.
- Fligner, M. A., and Rust, S. W. (1984). A modification of the Kruskal-Wallis statistic for the generalized Behrens-Fisher problem. *Communications in Statistics, Theory and Methods*, 13, 2013-2027.
- Hettmansperger, T. P. (1973). A large sample conservative test for the location with unknown scale parameters. *Journal of the American Statistical Association*, 68, 466-468.
- Hettmansperger, T. P., and Malin, J. S. (1975). A modified Mood's test for location with no shape assumptions on the underlying distributions. *Biometrika*, 62, 527-529.
- Hogg, R. V., Fisher, D. M., and Randles, R. H. (1975). A two-sample adaptive distribution-free test. *Journal of the American Statistical Association*, 70, 656-661.
- Kruskal, W. H. (1952). A nonparametric test for the several sample problem. *Annals of Mathematical Statistics*, 23, 525-540.
- Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47, 583-621.
- Mann, H. B., and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50-60.
- McNabb, D. E. (2004). *Research Methods for Political Science*. New York: Sharpe.
- Mood, A. M. (1954). On the asymptotic efficiency of certain non-parametric two-sample tests. *Annals of Mathematical Statistics*, 25, 514-522.
- Neubert, K., and Brunner, E. (2007). A studentized permutation test for the non-parametric Behrens-Fisher problem. *Computational Statistics and Data Analysis*, 51, 5192-5204.
- Potthoff, R. F. (1963). Use of the Wilcoxon statistic for the generalized Behrens-Fisher problem. *Annals of Mathematical Statistics*, 34, 1596-1599.

- Ramberg, J. S., and Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Communications of the Association of Computing Machinery*, 17, 78-82.
- Randles, R. H., Fligner, M. A., Policello, G. E., and Wolfe, D. A. (1980). An asymptotically distribution-free test for symmetry versus asymmetry. *Journal of the American Statistical Association*, 75, 168-172.
- Randles, R. H., and Wolfe, D. A. (1979). *Introduction to The Theory of Nonparametric Statistics*. New York: Wiley.
- Troendle, J. F. (2002). A likelihood ratio test for the nonparametric Behrens-Fisher problem. *Biometrical Journal*, 44, 813-824.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1, 80-83.
- Zaremba, S. K. (1962). A generalization of Wilcoxon's test. *Monatshefte für Mathematik*, 66, 359-370.

Authors' addresses:

Uttam Bandyopadhyay and Dhiman Dutta

Department of Statistics

University of Calcutta

35, Ballygunge Circular Road

Kolkata-700019, India

E-mail: ubandyopadhyay08@gmail.com and dhiman.dutta@gmail.com

Atanu Biswas

Applied Statistics Unit

Indian Statistical Institute

Kolkata, India

E-mail: atanu@isical.ac.in