Chapter 3

ADAPTIVE NONPARAMETRIC TESTS FOR TWO-SAMPLE LOCATION PROBLEM UNDER SYMMETRY

Abstract

Two adaptive test procedures are developed for the two-sample location problem under symmetry. The one having a deterministic approach is based on calculating a measure of tailweight, suggested by Crow and Siddiqui (1967), to select an appropriate set of rank scores for the two-sample linear rank statistic. The other having a probabilistic approach uses the p-value of an asymptotically distribution-free test for tailweight to select the appropriate set of scores. These test procedures are designed to have reasonably high power over a range of distributions. A simulation study is presented indicating that this procedure performs reasonably well in terms of power and attainment of the nominal size. The probabilistic approach is shown to be superior to the other existing competitors. Some relevant asymptotic properties are discussed. Both the adaptive procedures are illustrated by using a real data set.

3.1 Introduction

One of the fundamental problems in statistics, frequently encountered in applications, is the two-sample location problem. The observations within each sample are assumed to be independent and identically distributed. We further assume the independence between the two samples. In the location shift model the distribution functions are the same except perhaps for a change in their locations. This type of experiment arises frequently because the comparison of a standard treatment and a test treatment or an active compound versus a placebo is of interest. The design's simple set-up and execution also adds to its popularity.
The performance of any test statistic is dependent on the underlying distribution from which the data arise. The t-test is optimal for comparing the location parameters of two univariate normal populations given that the variances of the two populations are unknown but assumed equal. For heteroscedasticity and non-normal data, it would be more appropriate to apply a robust version of the t-test, like the Welch test or the trimmed t-test, or a nonparametric test. If the underlying population distributions are normal with unequal and unknown variances, either Welch's t-statistic or Satterthwaite's approximate F test can be suggested. However, Welch's procedure is non-robust under most non-normal distributions. The two-sample trimmed t-statistic according to Tukey and McLaughlin (1963) is one of the first papers on robust two-sample tests. Later Yuen and Dixon (1973) discussed this idea in more detail followed by a number of other works in this topic. Weichert and Hothorn (2002) developed robust hybrid tests for the two-sample location problem.

Nonparametric statistical procedures applicable to the two-sample location problem are based on the rank-order statistics for the combined samples, since various functions of these rank-order statistics can provide information about the possible difference between the populations. But the choice of a suitable rank test depends on the underlying distribution which may be symmetric or asymmetric and which may have light, medium or heavy tails. Usually a practicing statistician has no information about the underlying distribution of the data. Thus an adaptive test should be applied, taking into account the given data set. A two-sample adaptive distribution-free test was proposed by Hogg et al (1975). In this approach, the data was first used to assess the tailweight and skewness of the underlying distributions. Then an appropriate distribution-free test based on a simple linear rank statistic for shift in location was recommended. They showed that if the selected statistic were based on the order statistics and the tests were based on ranks, then the resulting test would maintain its level of significance. Further research on this interesting class of tests was successively
done by Ruberg (1986), Büning (1994) and O'Gorman (1997). Neuhäuser et al (2004) compared the several existing adaptive tests with a maximum test and a sum test, which are, respectively, given by the maximum and the sum of the absolute values of some standardized linear rank statistics for the two-sample location problem. A brief review on two-sample tests, including alternatives to Students t-test in non-normal models, is also presented by Reed (2005).

The purpose of this chapter is to provide two adaptive test procedures for the two-sample location problem under symmetry and to compare them with some non-parametric competitors. Since the performance of a nonparametric test based on linear rank statistic depends on the tailweight of the underlying distribution, it would therefore be desirable to choose the scores by using sample information about such population characteristic. The first adaptive procedure has a deterministic approach based on the concept of Hogg (1974). Here a function of the order statistics of the combined sample can be used as a selector statistic, and a measure based on quantiles suggested by Crow and Siddiqui (1967) is used to measure tailweight. Based on this measure, a distribution is classified as light or medium or heavy tailed. The choice of an appropriate set of scores then depends on the corresponding sample information. By using the value of the selector statistic to set the scores, the resulting adaptive test will no longer be a rank test. However, it will maintain the nonparametric distribution-free property as a rank test.

The second proposed adaptive test has a probabilistic approach. In the previous chapter we propose an adaptive test for the single sample location problem using the p-value from the triples test for symmetry given in Randles et al (1980). Here the proposed probabilistic approach uses the p-value of an asymptotically distribution-free test for tailweight. The test statistic for the preliminary test is based on the same measure as used in the deterministic approach. However, here the measure is calculated separately for each sample and then the test statistic for testing the tail-
weight is obtained by forming weighted average of these two-sample measures. The proposed probabilistic approach is shown to be asymptotically distribution-free and superior compared to the other existing competitors. Although the proposed test is constructed assuming the symmetry of the underlying distributions, it is asymptotically distribution-free even without the assumption of symmetry. The proposed test is shown in the simulation studies to be robust even for some asymmetric distributions and is only second to the Wilcoxon-Mann-Whitney test in terms of power.

The remaining part of the chapter is organized in the following way. Section 3.2 introduces the proposed adaptive test procedures. In Section 3.3 the proposed adaptive test procedures are illustrated using an example. Then in Section 3.4 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 3.5. Finally, Section 3.6 concludes with a discussion.

3.2 The Proposed Adaptive Tests

Let \( X_1, X_2, \ldots, X_{n_1} \) and \( Y_1, Y_2, \ldots, Y_{n_2} \) be random samples from populations with continuous distribution functions (d.f.'s) \( F(x) \) and \( F(x-\theta) \), \(-\infty < \theta < \infty\), respectively. Suppose \( F(x) + F(-x) = 1 \) for all \( x \). Then the objective is to test

\[
H_0 : \theta = 0
\]

against some composite alternative.

Under \( H_0 \), \( X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2} \) represent a single random sample of size \( N = n_1 + n_2 \) from a continuous symmetric d.f. \( F(x) \). A two-sample simple linear rank statistic is of the form

\[
A_T = \sum_{j=1}^{n_2} a_T(R_j),
\]

(3.2.2)
where \( R_j \) denotes the rank of \( Y_j \) among all \( N \) observations and \( a_T(1), a_T(2), \ldots, a_T(N) \) denote scores which satisfy nondecreasing and nonconstant conditions, i.e.,

\[
a_T(1) \leq a_T(2) \leq \ldots \leq a_T(N), \quad a_T(1) \neq a_T(N).
\]

For any such choice of scores \( A_T \) provides a nonparametric test for \( H_0 \). Also for each choice of the above scores we can find a best parametric test as efficient as the rank test in terms of local asymptotic power. But to select the appropriate rank test we must have knowledge about the underlying population. If the experimenter knew or could detect from the data the nature of the underlying distribution, he/she might be able to improve the power of the test procedure with a more appropriate choice of the scores \( a_T(1), a_T(2), \ldots, a_T(N) \). We therefore need to search among all rank tests for the one that is most powerful, in some sense, for a particular problem. Unfortunately, there is no such uniformly most powerful test procedure. So we seek the one that is locally most powerful.

First we consider the Wilcoxon-Mann-Whitney test \( A_W \) by letting

\[
a_W(i) = i, \quad \text{for } i = 1, 2, \ldots, N,
\]

in the linear rank statistic. This test is the locally most powerful rank test for detecting shift in logistic distribution, and it also has good power for detecting shifts in medium tailed symmetric distribution.

For a light tailed symmetric model, use the rank test \( A_L \) corresponding to scores given by

\[
a_L(i) = \begin{cases} 
  i - \lfloor (N + 1)/4 \rfloor & \text{if } i \leq (N + 1)/4 \\ 
  0 & \text{if } (N + 1)/4 < i < 3(N + 1)/4 \\ 
  i - \lceil 3(N + 1)/4 \rceil & \text{if } i \geq 3(N + 1)/4,
\end{cases}
\]

for \( i = 1, 2, \ldots, N \), where \( \lfloor \cdot \rfloor \) denotes the greatest integer less than or equal to \( \cdot \).
If the data indicate a model which is heavy tailed and symmetric, then use the rank test \((A_H)\) defined by the scores

\[
a_H(i) = \begin{cases} 
-[(N+1)/4] & \text{if } i < (N+1)/4 \\
\quad i - [(N+1)/2] & \text{if } (N+1)/4 < i < 3(N+1)/4 \\
\quad [(N+1)/4] & \text{if } i > 3(N+1)/4,
\end{cases}
\]

for \(i = 1, 2, \ldots, N\).

In practice the distribution is usually \textit{apriori} unknown and consequently one needs a test that has high relative power across the different possible distributions. We devote the rest of this section to introduce two adaptive test procedures which are designed for this aim.

3.2.1 The Deterministic Approach

We first propose an adaptive two-sample distribution-free test having a deterministic approach using a fairly easy classification scheme which merely attempts to detect the tailweight of the underlying symmetric distribution. To describe the classification scheme, let us first assume that \(\theta = 0\). Next let \(Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(N)}\) denote the order statistics of the combined sample of the \(N = n_1 + n_2\) observations \(X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2}\); i.e. these are the order statistics of a random sample of size \(N\) from \(F(x)\). We use the data to assess the tailweight of the underlying distributions. An interesting measure based on quantiles, used in Crow and Siddiqui, is given by

\[
TW = \frac{F^{-1}(1 - \beta_1) - F^{-1}(\beta_1)}{F^{-1}(1 - \beta_2) - F^{-1}(\beta_2)},
\]

where \(0 < \beta_1 < \beta_2 < 0.5\) with

\[
F^{-1}(u) = \inf \{x : F(x) \geq u\}, 0 < u < 1.
\]
Their choices for $\beta_1$ and $\beta_2$ are 0.025 and 0.25, respectively. Henceforth, we write $\hat{F}_n(x)$ to represent the empirical d.f. based on a sample of size $n$ and $\hat{F}_n^{-1}(u)$ to represent the corresponding quantile of order $u$, $0 < u < 1$. Then the statistic

$$\overline{TW} = \frac{\hat{F}_n^{-1}(0.975) - \hat{F}_n^{-1}(0.025)}{\hat{F}_n^{-1}(0.75) - \hat{F}_n^{-1}(0.25)},$$

is used to determine whether the tailweight of the underlying distribution is heavy or light.

Obviously $TW > 1$ and the longer tails represent the greater $TW$ values. In Table 3.1 values of $TW$ are presented for some selected symmetric distributions which are included in our simulation study in Section 3.4.

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Uniform</th>
<th>Normal</th>
<th>Logistic</th>
<th>Laplace</th>
<th>Cauchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TW$</td>
<td>1.9</td>
<td>2.91</td>
<td>3.33</td>
<td>4.32</td>
<td>12.71</td>
</tr>
</tbody>
</table>

The proposed deterministic approach (AD3) will accordingly use the following classification scheme. We compute $\overline{TW}$ using the combined sample of all $N$ observations. If $\overline{TW} < c_1$, we use a light tailed symmetric model, if $c_1 \leq \overline{TW} \leq c_2$, we decide to use a medium tailed symmetric model and finally if $\overline{TW} > c_2$, we use a heavy tailed symmetric model. We now propose the adaptive test AD3, combining the three linear rank statistics $A_L$, $A_W$ and $A_H$, in the following way:

$$AD3 = \begin{cases} 
A_L & \text{if } \overline{TW} < c_1 \\
A_W & \text{if } c_1 \leq \overline{TW} \leq c_2 \\
A_H & \text{if } \overline{TW} > c_2. 
\end{cases}$$

That means, if the data indicate a light tailed symmetric model we will use the L test. When the data indicate the medium tailed symmetric model we will use the
W test. If the unknown symmetric distribution is classified as heavy tailed we will apply the H test. Various choices of \( c_1 \) and \( c_2 \) are considered for the simulation study and no substantial deviation from the chosen nominal level of significance \( \alpha = 0.05 \) is observed for the AD3 test. However the power of the adaptive procedure differs for the different choices of \( c_1 \) and \( c_2 \). Hence, based on simulation study, we choose that \( c_1 \) and \( c_2 \) for which the AD3 test seems to be most powerful. As a result, \( c_1 = 2.2 \) and \( c_2 = 5.3 \) are found to be the best choice in terms of the power of the test.

To verify that the proposed deterministic approach is actually distribution-free, we note that the preliminary model selection was based on the tailweight statistic \( \bar{T} \bar{W} \), which is a function of the order statistics of combined sample of all \( N \) values. Under \( H_0 \), the rank statistics are each independent of the order statistics of the combined sample, and hence of \( \bar{T} \bar{W} \). Thus, under \( H_0 \), the preliminary selection of the model is independent of the final test. Since each of these tests is distribution-free the adaptive test is also distribution-free.

Also note that the preliminary selection of the model is based on the order statistics of the combined sample, its performance is changed by the amount of shift in the underlying distribution. Thus, as \( \theta \) increases, the selection of an appropriate model becomes less accurate. However Hogg et al (1975) argued that this does not pose a major problem. According to them, if the amount of shift is small, the selection of the model is not changed very much and if the amount of shift is large, it is fairly unimportant which model is selected, because each of the corresponding rank tests can easily detect the shift.

In order to avoid the dependence of the measure \( TW \) on the location alternatives, we may consider the use of an alternative selector statistic. We may calculate separate tailweight measures for each of the two samples. The tailweight measure for each sample does not depend upon the location alternatives. We then use a weighted average of the two tailweight measures as the selector statistic. The adaptive test
thus obtained is no longer exactly distribution-free (EDF), since the new selector statistic is only uncorrelated with the rank statistic. But we are able to frame an asymptotically distribution-free (ADF) test with the help of such weighted selector statistic.

3.2.2 The Probabilistic Approach

In the proposed deterministic approach the choice of $c_1$ and $c_2$ are predetermined subjectively based on simulation studies. To avoid this subjectivity we now suggest a test procedure which does not use any predetermined set of values. This approach, called probabilistic approach, is based on p-values calculated from some preliminary test on tailweight. In the probabilistic approach we again make use of the same measure of tailweight proposed by Crow and Siddiqui. We obtain the tailweight measures for the two samples as

$$TW_k = \frac{F^{-1}_{w_1}(0.975) - F^{-1}_{w_2}(0.025)}{F^{-1}_{w_1}(0.75) - F^{-1}_{w_2}(0.25)}, \quad k = 1, 2,$$

where $F^{-1}_{w_1}$ and $F^{-1}_{w_2}$ denote, respectively, the sample quantiles based on X and Y samples.

It is clear that this tailweight measure is invariant under location and scale transformation. Thus $TW_1$ and $TW_2$ have the same distribution. Now we conduct some preliminary tests by framing the following testing problems. The underlying population is light tailed against the alternative that it is medium or heavy tailed and that the underlying population is medium tailed against the alternative that it is heavy tailed. To test the null hypothesis of light tailed model we take $TW = 1.9$, the population TW value for the uniform distribution and consider the problem of testing

$$H_{01} : TW = 1.9$$
against

\[ H_{11} : TW > 1.9. \]

To test the null hypothesis regarding medium tailed model we take \( TW = 3.33 \), the population TW value for the logistic distribution and set the testing problem here as

\[ H_{02} : TW = 3.33 \]

against

\[ H_{12} : TW > 3.33. \]

In order to perform tests for the above problems, we have to find the distribution of the tailweight measure \( TW \). The exact distribution of \( TW \) can not be found analytically. But an asymptotic distribution of \( TW \) can be obtained by considering the following result. The result is virtually identical to Proposition 1 of Schmid and Trede (2003).

**Result 3.1** Suppose \( X_{(1)} < X_{(2)} < \ldots < X_{(n)} \) denote order statistics of a sample of size \( n \) from a distribution \( F \) having density \( f(x) \) continuous and positive in a neighborhood of the quantiles \( \xi_{p_1}, \xi_{p_2}, \xi_{p_3}, \text{ and } \xi_{p_4} \) with \( 0 < p_1 < p_2 < p_3 < p_4 < 1 \). Then, if \( \hat{\xi}_p = F_n^{-1}(p) \) represents the quantile of order \( p \) based on \( X_1, X_2, \ldots, X_n \), the asymptotic distribution of

\[ \sqrt{n} \left( \frac{\xi_{p_4} - \xi_{p_1}}{\xi_{p_3} - \xi_{p_2}} - \frac{\xi_{p_4} - \xi_{p_1}}{\xi_{p_3} - \xi_{p_2}} \right) \]  

(3.2.7)

is normal with mean zero and variance \( b'\Sigma b \), where

\[ b = \left( -\frac{1}{\xi_{p_3} - \xi_{p_4}}, \frac{\xi_{p_4} - \xi_{p_1}}{(\xi_{p_3} - \xi_{p_4})^2}, -\frac{\xi_{p_4} - \xi_{p_1}}{(\xi_{p_3} - \xi_{p_4})^2}, \frac{1}{\xi_{p_3} - \xi_{p_4}} \right) \]

and \( \Sigma \) is the \( 4 \times 4 \) symmetric matrix with \((i,j)\)th element

\[ \sigma_{ij} = p_i(1 - p_j)/f(\xi_{p_i})f(\xi_{p_j}), \quad i \leq j. \]
Proof. Let us consider the function

\[ g(x) = \frac{x_4 - x_1}{x_3 - x_2} \]

of 4-variables, having the partial derivatives

\[ \nabla g(x) = (g_1(x), g_2(x), g_3(x), g_4(x))', \]

where

\[ g_4(x) = -g_1(x) = -(x_3 - x_2)^{-1}, \]
\[ g_2(x) = -g_3(x) = (x_4 - x_1)(x_3 - x_2)^{-2}. \]

Then

\[ b = \nabla g(\xi_{p_1}, \xi_{p_2}, \xi_{p_3}, \xi_{p_4}) \]

and, by multivariate δ-method, the asymptotic distribution of (3.2.7) is same as that of \( b'\xi \), where

\[ \xi = \sqrt{n}(\hat{\xi}_{p_1} - \xi_{p_1}, \hat{\xi}_{p_2} - \xi_{p_2}, \hat{\xi}_{p_3} - \xi_{p_3}, \hat{\xi}_{p_4} - \xi_{p_4})'. \]

But, by our assumption, the asymptotic distribution of \( \xi \) is 4-variate normal with mean zero and variance-covariance matrix \( \Sigma \), and hence \( b'\xi \) is asymptotically normal with mean zero and variance \( b'\Sigma b \), which is our desired result. \( \square \)

From the above result we have \( \sigma_{TW}^2 \) such that the asymptotic distribution of

\[ \sqrt{n_k}(\hat{T}W_k - TW) \]

is normal with mean zero and variance \( \sigma_{TW}^2 \) as \( n_k \to \infty, k = 1, 2 \). Hence we find that

\[ \text{var}(\hat{T}W_k) \approx \frac{\sigma_{TW}^2}{n_k} = V_k, \text{ say.} \]
To obtain the test statistic for the preliminary tests we combine the two tailweight measures as

$$T_W = \frac{T_{W1} + T_{W2}}{N} = \frac{n_1}{N} T_{W1} + \frac{n_2}{N} T_{W2}.$$  

Since the measure $T_W$, used in the AD3 test, is computed on the basis of the combined sample observations it estimates the tailweight of a distribution which is not $F(x)$ but a mixture of $F(x)$ and $F(x - \theta)$. The sampling distribution of $T_W$ depends heavily on this fact. Hüsler (1987) observed that for heavy tailed distribution the bias and the standard deviation of $T_W$ is decreasing with increasing $\theta$. For light tailed distribution the mean of $T_W$ is increasing with $\theta$. These contrary effects induce the stable behavior of $T_W$ in the normal distribution. This dependence of the measure $T_W$ on the location alternatives has an influence on the power of the adaptive test procedures, since only a moderate rank test is selected by a misclassification instead of a rank test with good power. In order to avoid this dependence we use $T_{WC}$, which does not depend on the location alternatives, for the proposed probabilistic approach.

We now state the following result which gives the asymptotic normality of $T_{WC}$.

**Result 3.2** Suppose, as $\min(n_1, n_2) \to \infty$,

$$\frac{n_1}{N} \to \lambda, \quad \frac{n_2}{N} \to (1 - \lambda), \quad 0 < \lambda < 1.$$

Then the asymptotic distribution of

$$\sqrt{N}(T_{WC} - TW)$$

is normal with mean zero and variance $\sigma_{TW}^2$ as $\min(n_1, n_2) \to \infty$.

A natural choice of $\hat{\sigma}_{TW}$, a consistent estimator of $\sigma_{TW}$, is given by

$$\hat{\sigma}_{TW}^2 = \frac{n_1}{N} v_1 + \frac{n_2}{N} v_2,$$
where $v_k$ is obtained from $b'Bb$ of Result 3.1 by replacing $\xi_{p_k}$'s by the corresponding sample quantiles \( \hat{\xi}_{p_k}^{(n)} = \hat{F}_{n_k}^{-1}(p_k), k = 1, 2 \) and by estimating the density function appropriately.

The density estimator of $f(\xi_p)$ based on a sample of $n$ observations is given by

\[
\hat{f}_n(\xi_p) = \frac{\hat{F}_n(U_n) - \hat{F}_n(L_n)}{U_n - L_n},
\]

where we use the idea of Fligner and Rust (1982) with $L_n = X_{\lfloor np_0 + K_n\rfloor}$, $U_n = X_{\lfloor np_0 + 1 + K_n\rfloor}$ and $\hat{F}_n$ being the empirical d.f.. Result 3.4 in Section 3.5 gives sufficient conditions on the sequence $\{K_n\}$ to ensure that $\hat{f}_n$ converges in probability to $f$.

The preliminary test is thus based on the following two asymptotically normally distributed statistics

\[
U = \frac{\sqrt{N}}{\tau_{TW}} (\bar{T}_W - 1.9)
\]

and

\[
V = \frac{\sqrt{N}}{\tau_{TW}} (\bar{T}_W - 3.33).
\]

The corresponding test rules are then: Reject $H_0$ if $U > \tau_\alpha$ and reject $H_0$ if $V > \tau_\alpha$, where $\tau_\alpha$ is the upper $\alpha$th quantile of a standard normal distribution. That is the upper U-test and the upper V-test are appropriate for testing $H_0$ against $H_1$ and $H_0$ against $H_2$, respectively.

We now introduce the proposed adaptive rule AD4. Let $p_1^+$ denote the p-value corresponding to an observed $U$ for testing $H_{01}$ against $H_{11}$ i.e. $p_1^+ = P_{H_{01}}(U \geq u)$ and $p_2^+$ denote the p-value corresponding to an observed $V$ for testing $H_{02}$ against $H_{12}$ i.e. $p_2^+ = P_{H_{02}}(V \geq v)$. The selection rule should be defined on the basis of the classification probabilities $\pi_1 = p_1^+, \pi_2 = p_2^+(1 - p_1^+)$ and $\pi_3 = (1 - p_1^+)(1 - p_2^+)$. Whenever $p_1^+, p_2^+$ are observed, perform a random experiment having three possible outcomes with probabilities $\pi_1, \pi_2$ and $\pi_3$, where $\pi_1 + \pi_2 + \pi_3 = 1$. The adaptive rule is: Reject $H_0$ with probability $\pi_1$ if $A_L > A_L(\alpha, n_1, n_2)$, with probability $\pi_2$ if
$A_W > A_W(\alpha, n_1, n_2)$ and with probability $\pi_3$ if $A_H > A_H(\alpha, n_1, n_2)$. Equivalently we may say: Accept $H_0$ with probability $\pi_1$ if $A_L \leq A_L(\alpha, n_1, n_2)$, with probability $\pi_2$ if $A_W \leq A_W(\alpha, n_1, n_2)$ and with probability $\pi_3$ if $A_H \leq A_H(\alpha, n_1, n_2)$, where $A_L(\alpha, n_1, n_2)$, $A_W(\alpha, n_1, n_2)$ and $A_H(\alpha, n_1, n_2)$ are the upper $\alpha$-critical values for the L, W and H tests, respectively. Since the distributions of rank test statistics are discrete, we have to randomize each component test to exhaust the $\alpha$ level.

### 3.3 Example

As an example to illustrate the methods we consider data given in Bolton and Bon (2004). Two different formulations of a tablet of a new drug are to be compared with regard to rate of dissolution. Ten tablets of each formulation are tested, and the percent dissolution after 15 minutes in the dissolution apparatus is observed. The data is presented in the following table:

<table>
<thead>
<tr>
<th>Table 3.2</th>
<th>Percent dissolution after 15 minutes for two tablet formulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulation A</td>
<td>74, 71, 79, 63, 80, 61, 69, 72, 80, 65</td>
</tr>
<tr>
<td>Formulation B</td>
<td>68, 84, 81, 85, 75, 69, 80, 76, 79, 74</td>
</tr>
</tbody>
</table>

Figure 2 represents the boxplots of X and Y variables. From the boxplots it is clear that we may assume the underlying population to be symmetric.

For testing $H_0 : \theta = 0$ against the alternative $H_1 : \theta > 0$ we reject $H_0$ in favor of $H_1$ using the W test if and only if $A_W \geq A_W(\alpha, n_1, n_2)$. The observed value of $A_W = 128.5$ while $A_W(0.05, 10, 10) = 127$. Since observed $A_W > A_W(0.05, 10, 10)$, we would reject the null hypothesis $H_0$ with this data using the W test at 5% level of significance. Our level $\alpha = 0.05$ test based on the H test statistic is to reject $H_0$. 

56
in favor of $H_1$ if and only if $A_H > A_H(\alpha, n_1, n_2)$. The observed value of $A_H = 16.5$ is less than the tabulated value $A_H(0.05, 10, 10) = 18$. So based on this data we accept the null hypothesis using the H test. Using the L test with significance level $\alpha = 0.05$ we reject $H_0$ in favor of $H_1$ if and only if $A_L \geq A_L(\alpha, n_1, n_2)$. For this data observed $A_L = 12$ while $A_L(0.05, 10, 10) = 10$. So on the basis of the given data we conclude that we reject $H_0$ using the L test. The cut-off points of the rank tests are obtained from simulation study. Clearly there is difference in decision between the three tests based on linear rank statistics. So we may now proceed to illustrate the application of the two proposed adaptive procedures. We first perform the AD3 test. For this we calculate $TW = 2.052273$. The observed $TW$ is less than 2.2 and hence we use the L test and therefore $H_0$ is rejected at 5% level of significance as is seen previously.

To perform the AD4 test we need to compute the p-values for the preliminary tests. But before calculating the p-values we first need to calculate $T\bar{W}_1 = 1.578723$ and $T\bar{W}_2 = 2.546154$, and hence we obtain the combined tailweight measure as $T\bar{W}_c = 2.062439$. In this example we take $K_n = \frac{1}{2}n^\frac{1}{2}$. Thus we obtain $v_1 = 2.254569$, $v_2 = 11.81248$, and hence $\hat{\sigma}^2_{T\bar{W}} = 7.033524$. The observed values of $U$ and $V$
are, respectively, 0.2739163 and -2.137458. Then we calculate $p_u^+ = 0.3920745$ and 
$p_u^- = 0.9837196$, which give $\pi_1 = 0.3920745$, $\pi_2 = 0.5980282$ and $\pi_3 = 0.009897267$.

We now perform a random experiment having three possible outcomes $E_1$, $E_2$ and $E_3$
with probabilities 0.3920745, 0.5980282 and 0.009897267, respectively. If $E_1$ occurs we
use the L test, if $E_2$ occurs we use the W test and if $E_3$ occurs we use the H test. Suppose $E_2$
occurs then we use the W test and hence reject $H_0$ at 5% level of significance.

3.4 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 nonparametric
competitors W test, L test, H test, along with the median test (M) by taking the
median scores

$$a_M(i) = \begin{cases} 
1 & \text{if } i \geq (N + 1)/2 \\
0 & \text{if } i < (N + 1)/2.
\end{cases}$$

in the linear rank statistics given in (3.2.2) and the test based on normal scores (NS)
corresponding to the scores given by

$$a_{NS}(i) = \Phi^{-1}\left(\frac{i}{N+1}\right), \ i = 1, 2, \ldots, N.$$

The results are given for the upper tailed alternatives only. We perform simulations
for the following symmetric distributions:

(I) Uniform distribution (density with light tailweight): $U(0,1)$
(II) Normal distribution (density with medium tailweight): $N(0,1)$
(III) Logistic distribution (density with medium tailweight): $L(0,1)$
(IV) Double exponential distribution (density with heavy tailweight): $DE(0,1)$
(V) Cauchy distribution (density with very heavy tailweight): $C(0,1)$.
We also consider the following asymmetric distributions to assess the performance of the proposed tests when the assumption of symmetry is not satisfied:

(VI) Log-normal distribution with parameters $\mu = 0$ and $\sigma = 1$ (density with medium tailweight): LN(0,1)
(VI) Log-normal distribution with parameters $\mu = 0$ and $\sigma = 2$ (density with very heavy tailweight): LN(0,2).

The nominal significance level of the test is taken to be $\alpha = 0.05$ and 5000 simulation runs for each particular configuration. The empirical sizes and powers of the tests are computed as the proportions of cases in which the test statistics exceed the critical value. We investigate the powers at $\theta = \xi_{0.5}, \xi_{0.6}, \xi_{0.7}$, where $\xi_q$ is the $q$th quantile of the distribution of $X$ and the results are presented in Tables 3.3 and 3.4.

For the sample size of $n_1 = n_2 = 20$ there is no substantial deviation in actual type I error probability of the proposed adaptive procedures from the chosen nominal significance level $\alpha = 0.05$. For the uniform distribution, we observe that the proposed adaptive test AD4 is only second to the L test in terms of power comparison, well ahead of the proposed deterministic approach AD3. When the underlying distribution is normal, the proposed probabilistic approach AD4 is slightly anti-conservative and consequently more powerful compared to other competitors under study. In case of logistic distribution, the W test is the best test and both the AD3 and AD4 tests stay quite close to the W test. The M test emerges as the best test when the underlying distribution is double exponential while the AD4 test stays just ahead than the AD3 test, behind the H test and the W test. With the Cauchy distribution, the H test is the best test with the AD3 test and the M test staying quite close followed by the AD4 test. Now as far as the two asymmetric distributions are concerned the W test
Table 3.3
Empirical size and power of the tests for \( n_1 = n_2 = 20 \)

<table>
<thead>
<tr>
<th>q</th>
<th>W</th>
<th>M</th>
<th>L</th>
<th>H</th>
<th>NS</th>
<th>AD3</th>
<th>AD4</th>
</tr>
</thead>
<tbody>
<tr>
<td>U(0,1)</td>
<td>0.5</td>
<td>0.053</td>
<td>0.051</td>
<td>0.048</td>
<td>0.054</td>
<td>0.046</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.281</td>
<td>0.167</td>
<td>0.389</td>
<td>0.200</td>
<td>0.334</td>
<td>0.367</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.648</td>
<td>0.375</td>
<td>0.832</td>
<td>0.480</td>
<td>0.723</td>
<td>0.745</td>
</tr>
<tr>
<td>N(0,1)</td>
<td>0.5</td>
<td>0.050</td>
<td>0.053</td>
<td>0.053</td>
<td>0.047</td>
<td>0.046</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.194</td>
<td>0.161</td>
<td>0.171</td>
<td>0.183</td>
<td>0.194</td>
<td>0.191</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.485</td>
<td>0.372</td>
<td>0.430</td>
<td>0.446</td>
<td>0.486</td>
<td>0.477</td>
</tr>
<tr>
<td>L(0,1)</td>
<td>0.5</td>
<td>0.053</td>
<td>0.054</td>
<td>0.054</td>
<td>0.049</td>
<td>0.050</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.183</td>
<td>0.157</td>
<td>0.156</td>
<td>0.175</td>
<td>0.177</td>
<td>0.181</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.452</td>
<td>0.374</td>
<td>0.371</td>
<td>0.431</td>
<td>0.435</td>
<td>0.447</td>
</tr>
<tr>
<td>DE(0,1)</td>
<td>0.5</td>
<td>0.049</td>
<td>0.052</td>
<td>0.051</td>
<td>0.052</td>
<td>0.048</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.152</td>
<td>0.162</td>
<td>0.116</td>
<td>0.159</td>
<td>0.140</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.391</td>
<td>0.401</td>
<td>0.254</td>
<td>0.405</td>
<td>0.358</td>
<td>0.385</td>
</tr>
<tr>
<td>C(0,1)</td>
<td>0.5</td>
<td>0.050</td>
<td>0.054</td>
<td>0.051</td>
<td>0.050</td>
<td>0.046</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.143</td>
<td>0.154</td>
<td>0.093</td>
<td>0.158</td>
<td>0.123</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.343</td>
<td>0.385</td>
<td>0.165</td>
<td>0.388</td>
<td>0.276</td>
<td>0.383</td>
</tr>
<tr>
<td>LN(0,1)</td>
<td>0.5</td>
<td>0.048</td>
<td>0.052</td>
<td>0.048</td>
<td>0.048</td>
<td>0.054</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.193</td>
<td>0.154</td>
<td>0.175</td>
<td>0.181</td>
<td>0.190</td>
<td>0.190</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.478</td>
<td>0.355</td>
<td>0.435</td>
<td>0.439</td>
<td>0.481</td>
<td>0.470</td>
</tr>
<tr>
<td>LN(0,2)</td>
<td>0.5</td>
<td>0.053</td>
<td>0.050</td>
<td>0.049</td>
<td>0.049</td>
<td>0.047</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.199</td>
<td>0.147</td>
<td>0.172</td>
<td>0.180</td>
<td>0.193</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.480</td>
<td>0.373</td>
<td>0.435</td>
<td>0.440</td>
<td>0.483</td>
<td>0.442</td>
</tr>
</tbody>
</table>

emerges as the best test in both the situations. The AD3 and AD4 tests are found to maintain their nominal level very well and also the power of each of them is better than the competitors other than the W test.

Again, for the sample with \( n_1 = 25, n_2 = 15 \), both the proposed adaptive procedures produce empirical size close to the nominal level of significance. When the underlying distribution is uniform, the L test emerges as the best test among the
Table 3.4
Empirical size and power of the tests for $n_1 = 25, n_2 = 15$

<table>
<thead>
<tr>
<th>q</th>
<th>W M</th>
<th>L H</th>
<th>NS AD3</th>
<th>AD4</th>
</tr>
</thead>
<tbody>
<tr>
<td>U(0,1)</td>
<td>0.5</td>
<td>0.050 0.053</td>
<td>0.054 0.052</td>
<td>0.049 0.054</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.254 0.157</td>
<td>0.379 0.185</td>
<td>0.319 0.344</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.821 0.336</td>
<td>0.798 0.468</td>
<td>0.717 0.718</td>
</tr>
<tr>
<td>N(0,1)</td>
<td>0.5</td>
<td>0.053 0.048</td>
<td>0.051 0.051</td>
<td>0.052 0.049</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.191 0.148</td>
<td>0.176 0.173</td>
<td>0.193 0.186</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.459 0.341</td>
<td>0.431 0.422</td>
<td>0.473 0.449</td>
</tr>
<tr>
<td>L(0,1)</td>
<td>0.5</td>
<td>0.048 0.049</td>
<td>0.053 0.047</td>
<td>0.050 0.053</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.178 0.150</td>
<td>0.156 0.169</td>
<td>0.174 0.169</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.429 0.346</td>
<td>0.367 0.411</td>
<td>0.419 0.414</td>
</tr>
<tr>
<td>DE(0,1)</td>
<td>0.5</td>
<td>0.050 0.051</td>
<td>0.050 0.048</td>
<td>0.049 0.048</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.148 0.149</td>
<td>0.111 0.150</td>
<td>0.140 0.147</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.361 0.357</td>
<td>0.247 0.384</td>
<td>0.347 0.372</td>
</tr>
<tr>
<td>C(0,1)</td>
<td>0.5</td>
<td>0.053 0.050</td>
<td>0.049 0.052</td>
<td>0.050 0.050</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.140 0.146</td>
<td>0.092 0.151</td>
<td>0.126 0.148</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.320 0.350</td>
<td>0.169 0.371</td>
<td>0.278 0.364</td>
</tr>
<tr>
<td>LN(0,1)</td>
<td>0.5</td>
<td>0.051 0.050</td>
<td>0.047 0.047</td>
<td>0.047 0.048</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.185 0.142</td>
<td>0.171 0.172</td>
<td>0.165 0.177</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.465 0.331</td>
<td>0.421 0.419</td>
<td>0.444 0.439</td>
</tr>
<tr>
<td>LN(0,2)</td>
<td>0.5</td>
<td>0.052 0.050</td>
<td>0.059 0.047</td>
<td>0.049 0.046</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.184 0.144</td>
<td>0.177 0.173</td>
<td>0.172 0.177</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.465 0.335</td>
<td>0.423 0.420</td>
<td>0.446 0.422</td>
</tr>
</tbody>
</table>

competing procedures with the adaptive procedure AD4 at the second position slightly ahead of the AD3 test. The NS test is only slightly better than the W test when the distribution is normal, followed by the proposed probabilistic approach AD4 and ahead of the deterministic approach AD3. For logistic distribution there is no clear winner with respect to power, but the W test may have a very slight edge over the AD4 test which is reasonably better than the AD3 test. With the double exponen-
tial distribution, simulation study shows that the H test is the best test followed by the AD4 test. The H test is again the best test when the underlying distribution is Cauchy. Here the AD3 test gets the second position just ahead of the AD4 test. For the log-normal distributions the W test is again the best test with the proposed AD4 test in the second position in terms of the power. The M test however has very low power for the asymmetric densities considered here.

Discussion: From the results of the simulation study, we may conclude that both the proposed adaptive test procedures AD3 and AD4 are robust for nearly all cases. The proposed probabilistic approach AD4, although tends to be slightly anti-conservative in a few situations, is reasonably better than the AD3 test in terms of the total error. The adaptive tests are not the best one for a specific distribution but mostly second or third test. This is just the philosophy of an adaptive test to select the best one for a given data set. Note that the proposed adaptive test AD4 has a power which is only slightly below that of the W test, which is the best test, when the underlying distribution has a medium tailweight. In some situations the improvement of AD4 over AD3 is quite remarkable. When the underlying distribution is heavy tailed and symmetric the AD4 test remains competitive but, as expected, the H test performs a little better. In comparison to the other existing competitors the proposed adaptive tests perform very well in terms of powers. The M test also performs significantly well in this situation. If the underlying distribution is light tailed and symmetric, the increase in power of the AD4 test is quite satisfactory. The proposed adaptive test AD4 has the nearest power to the L test which is the best test in this situation. Thus the adaptive procedures are the most suitable among all tests considered. The AD3 test is exactly distribution-free and thus maintains the level of significance somewhat better than the AD4 test, which is only asymptotically distribution-free. Tables 3.3 and 3.4 show that, for the AD4 test, the empirical levels lie between 0.046 and 0.057,
which is acceptable. Therefore for the proposed adaptive tests the actual level can be
taken to be the same as the nominal level. The nominal level is maintained very well
even for the asymmetric distributions. Also their powers are always relatively close
to the powers of the best single test. Often, there is only a small difference in power
between the AD3 test and the AD4 test. The AD4 test is much more powerful than
the AD3 test in some situations, and hence the AD4 test seems to be preferable to
the AD3 test.

3.5 Some Asymptotic Properties

In this section we discuss some asymptotic properties of the proposed adaptive test
statistics. We know, as in Hájek et al. (1999), that the two-sample linear rank
statistics are asymptotically normal \((\mu_T, \sigma_T^2)\) under \(H_0\) with

\[
\mu_T = \frac{n_2}{N} \sum_{i=1}^{N} a_T(i)
\]

and

\[
\sigma_T^2 = \frac{n_1 n_2}{N(N-1)} \sum_{i=1}^{N} \left( a_T(i) - \frac{1}{N} \sum_{i=1}^{N} a_T(i) \right)^2
\]
as \(\min(n_1, n_2) \to \infty\), provided \(a_T(i)\)'s are generated by the relation

\[
a_T(i) = (N + 1) \phi_T \left( \frac{i}{N+1} \right), 1 \leq i \leq N,
\]

where \(\phi_T(u), 0 < u < 1\), is square integrable, non-constant and non-decreasing (or is a
difference between two monotonic functions). As a result, the asymptotic normality
under \(H_0\) of a properly normed two-sample linear rank statistic is ensured for the
common choices of scores. In the present case, if \(I(x)\) denotes the indicator function
assuming the value 1 or 0 according as \(x\) is true or false, the respective \(\phi_T\) functions,
defined on \((0, 1)\), are given by

\[
\phi_L(u) = \left( u - \frac{1}{4} \right) I \left( u \leq \frac{1}{4} \right) + \left( u - \frac{3}{4} \right) I \left( u \geq \frac{3}{4} \right), \quad (3.5.1)
\]
\( \phi_W(u) = u, \) \hspace{1cm} (3.5.2)
\( \phi_H(u) = \frac{1}{4} I \left( u < \frac{1}{4} \right) + \left( u - \frac{1}{2} \right) I \left( \frac{1}{4} \leq u \leq \frac{3}{4} \right) + \frac{1}{4} I \left( u > \frac{3}{4} \right). \) \hspace{1cm} (3.5.3)

It can be easily verified that the above \( \phi_T \)'s are all square integrable functions on \((0,1)\) and satisfy
\[
\int_0^1 (\phi_T(u) - \bar{\phi}_T)^2 du > 0,
\]
where \( \bar{\phi}_T = \frac{1}{T} \int \phi_T(u) du. \) Moreover \( \phi_T \)'s are monotonically non-decreasing on \((0,1)\), and hence by Lemma 1, p. 195, Hájek et al. (1999) it follows that
\[
\lim_{N \to \infty} \frac{1}{N+1} \left[ \phi_T \left( \frac{1 + [u,N]}{N+1} \right) - \phi_T(u) \right]^2 du = 0. \hspace{1cm} (3.5.4)
\]

Thus, for \( \min(n_1,n_2) \to \infty \), the asymptotic null distribution of
\[
\frac{A_T - \mu_T}{\sigma_T}
\]
is standard normal when \( T = W, L \) and \( H \).

Also note that the magnitudes of the p-values \( p_u^+ \) and \( p_v^+ \) indicate the nature of the tailweights of the underlying distributions. If the data present sufficient evidence of being light tailed then \( p_u^+ \) approaches \( p_L = I(TW < 1.9) + \frac{1}{2} I(TW = 1.9) \) while \( p_v^+ \) approaches \( 1 \). Consequently \( \pi_1 \) approaches \( p_L \) while \( \pi_2 \) and \( \pi_3 \) approach \( \bar{p}_L = \frac{1}{2} I(TW = 1.9) \) and \( 0 \), respectively. If the underlying distribution is medium tailed then \( p_u^+ \) approaches \( p_M = I(TW < 3.33) + \frac{1}{2} I(TW = 3.33) \) while \( p_v^+ \) approaches \( 0 \). In this case \( \pi_2 \) approaches \( p_M \) whereas \( \pi_1 \) and \( \pi_3 \) approach \( 0 \) and \( \bar{p}_M = \frac{1}{2} I(TW = 3.33) \), respectively. In case of heavy tailed data \( p_u^+ \) and \( p_v^+ \), both approach \( 0 \) and consequently \( \pi_3 \) approaches \( p_H = 1 \) with both \( \pi_1 \) and \( \pi_2 \) tending to \( 0 \).

Keeping the above facts in mind we consider the following standardized form of AD4 to study its asymptotic behavior:
\[
AD4^* = \frac{A_L - \mu_L}{\sigma_L} I(U^* < \pi_1) + \frac{A_W - \mu_W}{\sigma_W} I(\pi_1 \leq U^* \leq \pi_1 + \pi_2) + \frac{A_H - \mu_H}{\sigma_H} I(U^* > \pi_1 + \pi_2),
\]
where \( U^* \) is uniformly distributed over \((0,1)\) and is independent of \( \{X_1, X_2, \ldots, X_{n_1}, Y_1, Y_2, \ldots, Y_{n_2}\} \). We now proceed to verify the asymptotic normality of \( AD4^* \) under \( H_0 \) in the following result:

**Result 3.3** For \( \min(n_1, n_2) \to \infty \), the statistic \( AD4^* \) has asymptotically standard normal distribution under \( H_0 \).

**Proof.** Let \( \Psi_N(\tau) \) denote the distribution function corresponding to \( AD4^* \) under \( H_0 \). Further we denote the events \([A_l-\mu_l \leq \tau]\), \([A_w-\mu_w \leq \tau]\) and \([A_H-\mu_H \leq \tau]\) by \( A_{1N}, A_{2N} \) and \( A_{3N} \), respectively, and the events \([U^* < \pi_1], [\pi_1 \leq U^* \leq \pi_1 + \pi_2] \) and \([U^* > \pi_1 + \pi_2]\) by \( B_{1N}, B_{2N} \) and \( B_{3N} \), respectively. Then we can write

\[
\Psi_N(\tau) = P_{H_0}(AD4^* \leq \tau) = \sum_{k=1}^{3} P_{H_0}(A_{kN} \cap B_{kN}).
\] (3.5.5)

Let \( D_1, D_2 \) and \( D_3 \) be the sets of symmetric densities corresponding to light-tailed, medium-tailed and heavy-tailed distributions, respectively. Then

\[
\lim_{n_1, n_2 \to \infty} P_{H_0}(B_{kN}) = p_L, \text{ if } f \in D_1, k = 1 \\
\quad = \bar{p}_L, \text{ if } f \in D_1, k = 2 \\
\quad = p_M, \text{ if } f \in D_2, k = 2 \\
\quad = \bar{p}_M, \text{ if } f \in D_2, k = 3 \\
\quad = p_H, \text{ if } f \in D_3, k = 3 \\
\quad = 0 \text{ otherwise.}
\]

As the asymptotic null distribution of \( \sigma_T^{-1}(A_T - \mu_T) \) is standard normal, we have

\[
\lim_{n_1, n_2 \to \infty} P_{H_0}(A_{kN}) = \Phi(\tau),
\]
where $\Phi(\tau)$ is the standardized normal distribution function. Now, using the fact that, asymptotically, the rank statistics are distributed independently of $TWC$ under $H_0$, we get

$$\lim_{n_1,n_2 \to \infty} P_{H_0}(A_{kN} \cap B_{kN}) = \Phi(\tau)p_L \quad \text{if } f \in D_1, k = 1$$

$$= \Phi(\tau)p_L, \quad \text{if } f \in D_1, k = 2$$

$$= \Phi(\tau)p_M, \quad \text{if } f \in D_2, k = 2$$

$$= \Phi(\tau)p_M, \quad \text{if } f \in D_2, k = 3$$

$$= \Phi(\tau)p_H, \quad \text{if } f \in D_3, k = 3$$

$$= 0 \quad \text{otherwise.}$$

Combining the above, the required result follows immediately from (3.5.5). □

We can similarly establish the asymptotic normality of the standardized form of AD3 under the null hypothesis. Note that the above result also holds even without the assumption of symmetry. The asymptotic power properties of the adaptive test procedures depend on the criteria and the statistic used. The power of the adaptive test converges to the power of the best component. Thus under a sequence of Pitman’s local alternatives the power of the proposed adaptive test procedures converge to the power of the L test when the underlying distribution is light tailed and symmetric, to the power of the W test if the underlying distribution is medium tailed and symmetric and to that of the H test in case the underlying distribution is heavy tailed and symmetric. To obtain the asymptotic power of the proposed adaptive procedure, we assume that for the sequence of alternatives $\theta = \theta_N$, as $\min(n_1,n_2) \to \infty, \theta_N \to 0$, but

$$\left[ \frac{n_1n_2}{N} \Theta(f) \right] \theta_N^2 \to b^2, \quad 0 < b^2 < \infty, \quad (3.5.6)$$
where $\mathfrak{I}(f)$, called the Fisher's Information for a known density $f$, is given by

$$
\mathfrak{I}(f) = \int_0^1 \phi^2(u, f) \, du
$$

with

$$
\phi(u, f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}, \quad 0 < u < 1,
$$

assuming that the derivative, $f'(\cdot)$, of the density, exists at all but at most a countable number of values. Then under the sequence of local alternatives defined by (3.5.6), through (3.5.4), we can claim (see Hájek et al., 1999, p.267) that the asymptotic distribution of $\sigma_T^{-1}(A_T - \mu_T)$, when $T = L, W$ and $H$, is normal with mean $\mu_{b,T}$ and variance unity, where

$$
\mu_{b,T} = b\rho_T
$$

with

$$
\rho_T = \left[ \int_0^1 \phi_T(u)\phi(u, f) \, du \right] \cdot \left\{ \int_0^1 \phi^2(u, f) \, du \int_0^1 [\phi_T(u) - \bar{\phi}_T]^2 \, du \right\}^{-\frac{1}{2}}.
$$

Further, excepting the two boundary cases, viz., $TW = 1.9$ and $TW = 3.33$, we have

$$
\lim_{n_1,n_2 \to \infty} P(B_{kn}) = 1 \text{ or } 0 \quad (3.5.9)
$$

according as $f \in D_k$ or $f \notin D_k$. Hence, using (3.5.8), (3.5.9) and the inequality

$$
P(A_{kn}) + P(B_{kn}) - 1 \leq P(A_{kn} \cap B_{kn}) \leq \min(P(A_{kn}), P(B_{kn})), \quad (3.5.10)
$$

we find that

$$
\lim_{n_1,n_2 \to \infty} \Psi_N(\tau) = \Phi(\tau - b\rho^*)
$$

where

$$
\rho^* = \rho_L I(f \in D_1) + \rho_W I(f \in D_2) + \rho_H I(f \in D_3), \quad (3.5.11)
$$

excepting the cases mentioned above. Hence the asymptotic power of the upper level $\alpha$ AD4 test under (3.5.6) can be obtained as

$$
\beta(b) = \lim_{n_1,n_2 \to \infty} P(AD4^* > \tau_\alpha)
$$
where \( r_\alpha \) is the upper \( \alpha \)-quantile of the standard normal distribution. Similar expression can be obtained for the asymptotic power of the upper level \( \alpha \) AD3 test under (3.5.6). We next consider the sufficient conditions on the sequence \( \{K_n\} \) to ensure that \( \hat{f}_n \), given in (3.2.8), converges in probability to \( f \). The proof of the following result follows immediately using the same technique as in Fligner and Rust (1982). But for the sake of completeness we present here a brief outline of the proof.

**Result 3.4** Let \( X_1, X_2, \ldots, X_n \) be a random sample of size \( n \) from a population with distribution function \( F(x) \). Assume that \( F \) has at least two derivatives \( F' \) and \( F'' \) in some neighbourhood of \( \xi_p \). Further assume that \( F''(x) \) is bounded in this neighbourhood and that \( F'(\xi_p) = f(\xi_p) > 0 \). Then, if \( K_n = o(n^{1/2}\log n) \) and \( K_n(n^{1/4}\log n)^{-1} \to \infty \) as \( n \to \infty \), we have

\[
\hat{f}_n(\xi_p) \to f(\xi_p)
\]

in probability as \( n \to \infty \).

**Proof.** Let \( L_n = X_{[(np+1-K_n)]} \) and \( U_n = X_{[(np-K_n)]} \), and set

\[
\hat{f}_n(\xi_p) = \frac{\hat{F}_n(U_n) - \hat{F}_n(L_n)}{U_n - L_n},
\]

where \( \hat{F}_n \) is the empirical distribution function.

Suppose \( r = np \) is an integer. Using a standard inequality it can be easily shown that

\[
\lim_{n \to \infty} P(|U_n - \xi_p| > a_n) = \lim_{n \to \infty} P(|L_n - \xi_p| > a_n) = 0
\]

where \( a_n \sim n^{-1/2}\log n \).

Then, using Bahadur representation on quantiles, we have

\[
\hat{f}_n(\xi_p) = \frac{\hat{F}_n(U_n) - \hat{F}_n(L_n)}{U_n - L_n} = f(\xi_p)(U_n - L_n) + O_p(n^{-3/4}\log n)
\]

(3.5.12)
as $n \to \infty$. A similar application of Bahadur's representation also yields

$$2[K_n] = n\{\hat{F}_n(U_n) - \hat{F}_n(L_n)\} = nf(f_p)(U_n - L_n) + O_p(n^{1/4}\log n). \quad (3.5.13)$$

Equation (3.5.13) combined with the assumption $K_n(n^{1/4}\log n)^{-1} \to \infty$ as $n \to \infty$ shows that $n^{3/4}(\log n)^{-1}(U_n - L_n) \to \infty$ in probability as $n \to \infty$. This again combined with equation (3.5.12) gives our required result.

\[ \square \]

3.6 Concluding Remarks

In this chapter we have developed two adaptive test procedures for two-sample location problem under symmetry. The proposed deterministic approach AD3 maintains the designated $\alpha$ level fairly accurately while displaying good power properties. But the choice of the points $c_1$ and $c_2$ induces some subjectivity in the test procedure. However there is no such points $c_1, c_2$ in the proposed probabilistic approach. It only takes into account the p-values of the tests based on $TW_c$ for testing $H_{01}$ against $H_{11}$ and $H_{02}$ against $H_{12}$, and hence there is no such subjectivity in the test procedure. The deterministic approach based on the tailweight measure from the combined sample is distribution-free since $TW$ is independent of the rank tests. The measure $\overline{TW}$, however, is affected by the amount of shift under $H_1$. We may also consider an adaptive deterministic procedure based on a combined measure from the single samples. But this test would not be exactly distribution-free since, for example, $\overline{TW_c}$ in our case is only uncorrelated with the rank tests. However, asymptotically $\overline{TW_c}$ is also independent of the rank tests and thus the test would be only asymptotically distribution-free.

For the probabilistic approach various other choices of p-values are examined e.g., we may consider the problem of testing $H_{01}$ against $H_{11}$ along with the problem of
testing for the heavy tailed model. We may take $TW = 4.32$, the population TW value for the Laplace distribution, and set the testing problem as

$$H_{03} : TW = 4.32$$

against

$$H_{13} : TW < 4.32.$$

Then we have the corresponding p-values for these two tests as $p_u^+ = P_{H_0}(U \geq u)$ and $p_d^- = P_{H_0}(D \leq d)$ with $\pi_1 = p_u^+, \pi_2 = (1 - p_u^+)(1 - p_d^-)$ and $\pi_3 = p_d^- (1 - p_u^+)$ or with $\pi_1 = p_u^+(1 - p_d^-), \pi_2 = (1 - p_u^+)(1 - p_d^-)$ and $\pi_3 = p_d^-$, where

$$D = \frac{\sqrt{N}}{\hat{\sigma}_{TW}} (TW - 4.32).$$

For all these choices the adaptive test attains the nominal level $\alpha$ but the choice of classification probabilities, given in the subsection 3.2.2, seems to be the best for the proposed adaptive procedure $AD_4$ in terms of power of the test.

The adaptive methods described here are motivated by asymptotic properties. To investigate whether these methods can be used for small or moderate sample sizes, simulation studies are carried out and the relative performance of the proposed adaptive procedures with the existing nonadaptive competitors is assessed. The simulation studies in Section 3.4 show that the sizes of the proposed adaptive tests are near the nominal value and thus the proposed adaptive tests are valid tests for the two-sample location problem under symmetry. The simulations also show that the powers of the adaptive tests often nearly equaled the power of the best single test. It is also shown in the simulation studies that the proposed adaptive tests are robust and performs reasonably well in terms of power even when the underlying distribution is asymmetric. The proposed adaptive tests are not designed to be optimal for any particular distribution but this study convinces us that these adaptive tests are certainly worth considering in practical problems. In fact the proposed probabilistic approach $AD_4$
turns out to be the second best in most of the situations. So when nothing is known about the tailweight of a distribution the proposed probabilistic approach AD4 should be used for the two-sample location problem under symmetry, since considering only the size and power of the tests, there is little to lose and much to gain in using the adaptive procedures instead of the existing nonparametric competitors. Our example shows that the adaptive procedures are practical and reasonable.

It is always possible to increase the number of rank tests used in the adaptive scheme by making further classifications of the underlying distribution on the basis of the data at hand. However, unless a substantial increase in power is observed with the additional test, it is not worthwhile to create a new category. The additional test will tend to decrease the power of the adaptive test for detecting a shift in a distribution for which it is not as good as the other rank tests used in the adaptive scheme.

Some measures of tailweight for the distribution \( F \) are considered as selector statistics. A decision is to be made between measures of integral type (Hogg 1972, 1974, Hogg et al., 1975) or of quantile type. To investigate the tailweight of the underlying distribution Hogg used the following integral type measure

\[
\frac{U_{0.05} - L_{0.05}}{U_{0.5} - L_{0.5}},
\]

where \( U_{\beta} (L_{\beta}) \) is the average of the largest (smallest) \( \beta \) percent of the observations defined as:

\[
U_{\beta} = \frac{1}{\beta} \int_{F^{-1}(\beta)}^{1} F^{-1}(y)dy \quad (L_{\beta} = \frac{1}{\beta} \int_{0}^{F^{-1}(\beta)} F^{-1}(y)dy)
\]

for \( \beta \in (0, 1) \). According to Hogg's simulation studies \( \beta_1 = 0.05 \) and \( \beta_2 = 0.5 \) turns out to be the most satisfactory. Now in order to compare between quantile type measures used in our proposed adaptive procedure and integral type measures due to Hogg we have to take into account that quantile estimates become worse near the bounds of \([0, 1]\). The error of quantile estimation affects the integral type estimates more than the quantile type estimates. Another point in favour of quantile type measures is that they can
be computed more easily. Regarding the tailweight measure $TW$ based on quantiles, various choices of $\beta_1$ and $\beta_2$ have been used in a variety of contexts. Looking for a measure of tailweight one is tempted to make $\beta_1$ and $\beta_2$ small in order to capture the tail behavior of the distribution. On the other hand the difference between $\beta_1$ and $\beta_2$ must not be too large in order to remain within the tail of the distribution. Thus, on one hand one tries to avoid the application of extreme quantiles. On the other hand the quantiles should not be near the center of the distribution. Andrews et al. (1972) used $\beta_1 = 0.01$ and $\beta_2 = 0.25$ in (3.2.6) as an index of non-normality to assess the distribution of the estimators included in the Princeton Robustness Study. Schmid and Trede (2003) considered $\beta_1 = 0.125$, $\beta_2 = 0.25$ and $\beta_1 = 0.025$, $\beta_2 = 0.125$ in (3.2.6) for symmetric distributions. Crow and Siddiqui (1967) used $\beta_1 = 0.025$ and $\beta_2 = 0.25$ to rank, in order of increasing tail thickness, the symmetric distributions included in a comparative study of location estimators. Consequently, there is no choice of $\beta_1$ and $\beta_2$ which is fully satisfactory. As we have to take a decision among these various alternatives, the choice due to Crow and Siddiqui (1967) seems more preferable for our situation according to simulation studies.

Some relevant asymptotic properties of the proposed adaptive test procedures are discussed in Section 3.5. The asymptotic null distribution of the standardized form of the proposed adaptive test statistics are standard normal. The power of the proposed probabilistic approach converges to the asymptotic power of the best component under a sequence of local alternatives excepting the two boundary cases. In the next chapter we try to modify the probabilistic approach so that the asymptotic power of the adaptive test is equal to that of the best test at all points. We also focus our attention to another shortcoming of the AD4 test and suggest a method to overcome this drawback.