A DISTRIBUTION-FREE RANK-LIKE TEST FOR SCALE WITH UNEQUAL POPULATION LOCATIONS

by

R. Clifford Blair
University of South Florida

G.L. Thompson
Southern Methodist University

Technical Report No. SMU/DS/TR-248
Department of Statistical Science

Research Sponsored by NSF
Contract No. DMS 8918318

September 1990

A Distribution-Free Rank-Like Test For Scale With Unequal Population Locations

R. Clifford Blair
Department of Pediatrics
and
Department of Epidemiology and Biostatistics
University of South Florida
Tampa, Florida 33620

and

G. L. Thompson[1]
Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275

ABSTRACT

The properties of a distribution-free rank-like test for the two-sample scale problem are studied. This rank-like test is superior to commonly used rank tests for scale in that it: (1) does not require equal or known location parameters, (2) is robust for skewed data, (3) is resolving and (4) has significant power advantages in some circumstances. It is a statistic especially well suited for testing for equality of scale in biological applications where data are sampled from skewed populations with unequal medians. The proposed test is shown to be asymptotically normal and asymptotic relative efficiencies are calculated. Power properties are studied via simulation. Extensions to the j-sample problem are indicated.

[1] This research was supported in part by NSF grant DMS 8918318

1. Introduction

While powerful, robust, and highly useful distribution-free rank tests for differences in population locations are readily available, the same can not be said of tests for scale differences. The most common distribution-free tests for scale include the Siegel-Tukey test (S-T), the Ansari-Bradley test (A-B), the Capon test, Mood's test, and the Klotz test, all of which are linear rank statistics based on score functions that are symmetric (or nearly symmetric) about 1/2. Such distribution-free rank tests for scale are hampered by restrictive assumptions that significantly limit their applicability. One of the most common and problematic assumptions is that the two distributions do not differ in terms of a specified location parameter, typically the median. Or if the location parameters are unequal, it is assumed that they are at least known (Bradley, 1968). In this connection Moses (1963) stated "No rank test ... can hope to be a satisfactory test against dispersion alternatives without some sort of strong restrictions (e.g. equal or known medians) being placed on the class of admissible distribution pairs." Assumptions of this type are rarely tenable in the biological sciences and violations of such assumptions may entail serious statistical consequences. One of the most serious consequences of violating the assumption of equal medians in the commonly used rank tests for scale occurs with skewed data. When the data is not from a symmetric distribution, linear rank statistics for scale detect differences in location in the abscence of differences in scale. In addition to having restrictive assumptions and being nonrobust to unequal medians with skewed data, linear rank statistics for scale are nonresolving (Wasserstein and Boyer (1990)); that is, as the scale parameters of the two distributions move further apart, the probability of detecting the alternative does not approach 1.

The primary purpose of this paper is to examine the properties of a rank-like test for scale differences that (1) makes no assumptions concerning distribution locations, (2) is resolving, (3) does not detect differences in location for skewed distributions when the null hypothesis of equal scale parameters is true, (4) is distribution-free and (5) is referenced to readily available tables of critical values. This test is especially suitable for testing for scale differences when the two samples are from

skewed populations with different medians. Data of this sort are quite common in biological applications. In the remainder of this paper, a description of the test statistic is given, its asymptotic properties are discussed, its power properties are compared to that of the well known Siegel-Tukey test (Siegel and Tukey, 1960), and finally a brief discussion is provided.

2. Definition of the Test Statitsic.

Let X and Y be two random variables with distribution functions $F_X(t)$ and $F_Y(t)=F_X((t-\mu)/\sigma)$ and with densities $f_X(t)$ and $f_Y(t)=\frac{1}{\sigma}f_X((t-\mu)/\sigma)$, respectively. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ denote two random samples of sizes m and n from populations with cdfs $F_X(t)$ and $F_Y(t)$, respectively. Testing the null hypothesis that the two random variables, X and Y, have the same scale parameters is equivalent to testing the null hypotheses $H_o:\sigma=1$ versus the alternative $H_a:\sigma\neq 1$. No assumptions are made about the symmetry of $F_X(t)$ or about the equality of the location parameters of X and Y.

The proposed test statistic is constructed by first randomly choosing two observations, say X_i and X_j, from the sample $X_1, X_2, \ldots, X_m$ and defining the random variable $D_{x1}=|X_i-X_j|$. Then X_i and X_j are deleted from the sample and the procedure is repeated successively to produce $D_{x2}, \ldots, D_{x[m/2]}$ (where [x] denotes the greatest integer part of x). In generating $D_{x1}, \ldots, D_{x[m/2]}$, all the observations in the sample $X_1, X_2, \ldots, X_m$ are exhausted if m is even, or one observation remains if m is odd. In the same manner the random variables $D_{y1}, D_{y2}, \ldots, D_{y[n/2]}$ are generated from $Y_1, Y_2, \ldots, Y_n$. While the random variables D_{xi} and D_{yi} may seem at first glance unmotivated, it follows from Theorems 2 and 3 of Bickel and Lehmann (1979) that it is very natural and entirely expected that a test statistic for scale should be based on the random variables X_i-X_j and Y_i-Y_j if it is desired that the test statistic also be insensitive to location differences.

To study the distribution of the resulting random variables, let D_X and D_Y denote random variables with cdfs

$$F_{D_x}(t)=\int_{-\infty}^{\infty}\int_{u-t}^{t+u} f_x(v)\, f_x(u)\, dv\, du \quad , \quad t\geq 0,$$

$$\text{and} \quad F_{D_y}(t)=\int_{-\infty}^{\infty}\int_{u-t}^{t+u} f_y(v)\, f_y(u)\, dv\, du = F_{D_x}(t/\sigma) \quad , \quad t\geq 0,$$

respectively. Then, the generated random variables, $D_{x1}, D_{x2}, \ldots, D_{x[m/2]}$ and $D_{y1}, \ldots, D_{y[n/2]}$, are simply random samples with distribution functions $F_{D_x}(t)$ and $F_{D_y}(t)$. The densities for D_x , $f_{D_x}(t)$, are given in Table 1 for several common densities $f_x(t)$. Note that $F_{D_x}(t)$ and $F_{D_y}(t)$ are asymmetrical distributions in which the total mass is confined to the positive axis. Because the difference between $F_{D_x}(t)$ and $F_{D_y}(t)$ is completely described by the scale parameter, σ, the null hypothesis that X and Y have equal scale parameters can be tested without interference from the unequal (nuisance) location parameters simply by comparing the two random samples $D_{x1}, D_{x2}, \ldots, D_{x[m/2]}$ and $D_{y1}, \ldots, D_{y[n/2]}$. The proposed test statistic is motivated by noting that although there are no nuisance parameters to describe location differences between D_x and D_y, the medians of $F_{D_x}(t)$ and $F_{D_y}(t)$ are different if $\sigma\neq 1$. As σ moves further from 1 , the medians of D_x and D_y move further apart. Hence, the heuristic idea behind the proposed statistic for testing $H_o:\sigma=1$ versus $H_a:\sigma\neq 1$ is to apply any suitable two-sample rank test for location, such as the Wilcoxon rank-sum test, to the two random samples, $D_{x1}, D_{x2}, \ldots, D_{x[m/2]}$ and $D_{y1}, \ldots, D_{y[n/2]}$.

To formally define the proposed test, let R_{xi} and R_{yj} , $1\leq i\leq[m/2]$, $1\leq j\leq[n/2]$, denote the ranks of D_{xi} and D_{yj}, respectively, in the combined sample $D_{x1}, \ldots, D_{x[m/2]}, D_{y1}, \ldots, D_{y[n/2]}$. Let $N=[m/2]+[n/2]$, and let $\phi(t)$ be any square integrable function on (0,1) called the score function. Although this is a test for scale, we will be primarily interested in score functions that are monotone so as to detect the shift in medians that accompanies the change in scale between D_x and D_y. The scored ranks are generated from the score function either as approximate scores, $a(i)=\phi(i/N)$, or as exact scores, $a(i)=E(U_{(i)})$ where $U_{(i)}$ denotes the ith order statistic of a sample of size N from a uniform distribution on (0,1). Define the scored rank associated with D_{xi} as $a_{xi}=a(R_{xi})$. Then the proposed

test is the sum of the scored ranks associated with the [m/2] random vaiables $D_{x1}, D_{x2}, \ldots, D_{x[m/2]}$:

$$B=\sum_{i=1}^{[m/2]} a_{xi}.$$

For any choice of score function, ϕ, the linear rank statistic B is distribution-free. As breifly noted by Moses (1963), when ϕ is the identity function with approximate scores (that is, $\phi(t)=t$ and $a_{xi}=R_{xi}(N+1)^{-1}$), then (N+1)B is exactly the Wilcoxon rank-sum test applied to the two samples $D_{x1}, \ldots, D_{x[m/2]}$ and $D_{y1}, \ldots, D_{y[n/2]}$, and tables of small sample critical values are readily available. Duran and Mielke (1968) discuss the use of the score function $\phi(t)=\left(t(N+1)/N\right)^2$ for comparing the scale parameters of two asymmetrical one-sided distributions, and Mielke extends the discussion to score functions $\phi(t)=\left(t(N+1)/N\right)^r$, $r>0$. In this paper we will examine the test statistic B with two different sets of scores. First the Wilcoxon scores, which are widely used to test for differences in location, generate the test statistic

$$B_W=\sum_{i=1}^{[m/2]} R_{xi}/(N+1) .$$

On the other hand, the shape of f_{D_x} suggests that the Savage scores, $a_s(i)=\sum_{j=N=1-i}^{N}(1/j)$ may be appropriate, thus generating the test statistic

$$B_S=\sum_{i=1}^{[m/2]} a_s(R_{xi}) .$$

Also, locally most powerful tests for scale can be derived via exact scores. Define the function

$$\phi(u,f_{D_x})=-1-F^{-1}_{D_x}(u)\,\frac{f'_{D_x}(F^{-1}_{D_x}(u))}{f_{D_x}(F^{-1}_{D_x}(u))}$$

for $u\in(0,1)$. From Hajek and Sidak, Section VII.1.3 (1967), it follows that $\phi(u,f_{D_x})$ is the optimal score function and that the locally most powerful test for detecting changes in scale is

$$B_{\phi(u,f_{D_x})}=\sum_{i=1}^{[m/2]} E[\phi(U_{(i)},f_{D_x})] .$$

Note that B_S is the locally most powerful test for detecting changes in shift when the underlying data is exponential.

It is interesting to examine the proposed test, B, in light of the statement by Moses (1963) (as quoted in the Introduction of this paper). Previous attempts to construct rank-like tests for scale differences between random variables with unequal medians, such as Fligner and Killeen (1976), were based on transforming the two samples in an attempt to match up the medians. Unlike B, many of the resulting tests assume symmetric distribution functions. The test B is unique in that the two samples are transformed to have an entirely different location property in common. The distributions, $F_{D_x}(t)$ and $F_{D_y}(t)$, do not have equal medians, but they share the location property that $F_{D_x}(t)=F_{D_y}(t)=0$ for all $t\leq 0$ and for all values of σ. In many commonly encountered situations, the assumption of equal medians is unrealistic, but the assumption of equal endpoints of the supports is always satisfied by the construction of the random variables D_x and D_y . Hence, in testing whether $D_{x1}, D_{x2}, \ldots, D_{x[m/2]}$ and $D_{y1}, \ldots, D_{y[n/2]}$ are from populations with the same scale parameter, the usual restriction of equal medians is replaced by the (already satisfied) restriction of equal endpoints for the half-lines that describe the support of the two underlying populations from which D_x and D_y are sampled.

The proposed test, B, also posesses another desirable property not shared by many other linear rank tests for the equality of scale parameters, namely that the power of the test approaches 1 as the parameter of interest moves away from the null hypothesis and appoaches the extremes of the alternative hypothesis. When the null hypothesis is $H_0:\theta\in\omega$ and the alternative hypothesis is $H_a:\theta\in\Omega\backslash\omega$, Jogdeo (1966) defines a test to be <u>resolving</u> if $\sup_{\theta\in\Omega\backslash\omega}\Pr(\text{reject } H_0|\ \theta)=1$. Otherwise, it is said to be <u>nonresolving</u>. Wasserstein and Boyer (1990) show that a large class of linear rank tests for equality of scale are nonresolving. Included in the class of nonresolving tests for scale are the Ansari-Bradley test, the Capon test, the Siegel-Tukey test, the Klotz test, and the Mood test.

To show that B is resolving when $\theta=\sigma$ is the parameter of interest and when ϕ is nondecreasing, first note that $\Omega=(0,1)$ and $\omega=\{1\}$. The extremes in the alternative correspond either to when σ moves toward infinity, at which time all the observations in the sample $D_{x1}, D_{x2}, \ldots, D_{x[m/2]}$ will tend to preceed all of the observations $D_{y1}, D_{y2}, \ldots, D_{y[n/2]}$, or when σ moves to 0, in which case all of the observations in the sample $D_{x1}, D_{x2}, \ldots, D_{x[m/2]}$ will tend to follow all of the observations $D_{y1}, D_{y2}, \ldots, D_{y[n/2]}$.

Theorem 2.1. Assume that ϕ is nondecreasing. Then test statistic B is resolving and

$$\lim_{\sigma\to\infty}\Pr(B \text{ reject } H_0|\ \sigma)=\lim_{\sigma\to 0}\Pr(B \text{ reject } H_0|\ \sigma)=1\ .$$

Linear rank statistics for the two-sample location problem are generally resolving, while linear rank statistics for the two-sample scale problem generally are not. In this respect, the structure of B_N is more similar to the structure of rank tests for location than to that of rank tests for scale.

These ideas can be generalized to yield test statistics for the J-sample problem. Suppose there are J samples, each of size n_j from populations with cdfs $F((t-\mu_j)/\sigma_j)$, $1<j<J$. The null hypothesis is $H_0:\sigma_1=\sigma_2=\ldots=\sigma_J$ versus the alternative that not all of the σ_j , $1<j<J$, are equal. From each of the J samples, the corresponding samples, $D_{j1}, D_{j2}, \ldots, D_{j[n_j/2]}$, are generated as above. Then the Kruskall-Wallis test or another appropriate rank test for the J-sample location problem is applied to the samples of D_j's. The result is a distribution-free test for the J-sample scale problem that does not require equality of medians, is resolving, and is especially appropriate for skewed data.

3. Asymptotic Properties of B.

To study the asymptotic properties of B, first define the following:

$\lambda=\lim_{m\to\infty}[m/2]/N,$

$\mu_0=[m/2]/N\sum_{i=1}^{N} a(i),$ and

$$\sigma_0^2=[m/2][n/2]((N(N-1))^{-1}\sum_{i=1}^{N}\left(a(i)-N^{-1}\sum_{i=1}^{N}a(i)\right)^2 .$$

Assume that $0<\lambda<1$. Then, it follows (see Randles and Wolfe (1979) or Hajek and Sidak (1967)) that $(B-\mu_0)/\sigma_0$ converges in distribution to a standard normal random variable under the null hypothesis. This yields an easily implemanted large sample test for $H_0:\sigma=1$. For B_W straightforward computations show that $\mu_0=\frac{1}{2}[m/2]$, and $\sigma_0^2=[m/2]\times[n/2]/(12(N+1))$.

The asymptotic relative efficiency (ARE) of B relative to other tests for scale is obtained by by first calculating the efficacy of B for a sequence of Pitman alternatives for scale. It follows from Hajek and Sidak (1967) (cf also Randles and Wolfe (1979)) that the efficacy of B for any square intregrable score function is

$$\text{eff}(B) = \lambda(1-\lambda)\left(\int_{-\infty}^{\infty}\phi(u)\,\phi(u,f_{D_X})\,du\right)^2\left(\int_{-\infty}^{\infty}(\phi(u)-\overline{\phi})^2\,du\right)^{-1}$$

where $\overline{\phi}=\int_0^1\phi(u)du$. The expression for eff(B) can be simplified. Assuming mild regularity conditions on the underlying distribution, the first integral becomes

$$\int_{-\infty}^{\infty}\phi(u)\,\phi(u,f_{D_X})\,du = \int_{-\infty}^{\infty}u\phi'(F_{D_X}(u))\,f_{D_X}^2(u)\,du .$$

When ϕ is the identity function, we have $\overline{\phi}=\frac{1}{2}$ and (cf. Duran and Mielke (1968))

$$\text{eff}(B)=12\lambda(1-\lambda)\left(\int_0^{\infty}u\,f_{D_X}^2(u)\,du\right)^2$$

as long as $\lim_{t\to\infty}tf_{D_X}(t)=0$. Because the construction of B effectively halves the sample size, the ARE of B relative to any other test, say T, is

$$\text{ARE}(B,T)=\frac{\text{eff}(B)}{2\,\text{eff}(T)}.$$

The values for ARE(B_W,S-T) are given in Table 2 for a variety of distribution functions. The values for eff(S-T) are calculated by (cf Randles and Wolfe (1979)

$$\text{eff}(S\text{-}T)=48\lambda(1-\lambda)\left(\int_{\nu}^{\infty}u\,f_X^2(u)\,du-\int_{-\infty}^{\nu}u\,f_X^2(u)\,du\right)^2$$

where ν is the median of f_X . Because the Siegel-Tukey test and the Ansari-Bradley test are asymptotically equivalent, it follows that ARE(B_W,S-T)=ARE(B_W,A-B). Note that the ARE's increase as the tail weight of the underlying distribution incraeeases. In examining the ARE's in Table 2 it should be emphasized that these ARE's apply only in the case when the medians of the two samples are equal. In this case, however, B would not be the test of choice because the effective sample size is halved (and hence the power lowered) to compensate for the unequal medians. When the two medians are unequal, the Siegel-Tukey test, the Ansari-Bradley test, and the Mood test are not valid, but B is valid. Power comparisons in the case of unequal medians is explored in Section 4.

It is of particular interest to further examine the asymptotic properties of the more commonly used rank tests for scale (e.g. S-T, A-B, etc.) under the null hypothesis of equal scale parameters when the two samples have unequal medians. This can be accomplished by letting L be any linear rank statistic and examining the efficacy of L under a sequence of Pitman location alternatives. This efficacy is:

$$\mathrm{eff}(L) = \lambda(1-\lambda)\left(\int_{-\infty}^{\infty} \phi'(F_X(u))\, f_X^2(u)\, du\right)^2 \left(\int_{-\infty}^{\infty} (\phi(u)-\overline{\phi})^2\, du\right)^{-1} .$$

This expression is 0 if and only if the first integral is 0. When F is symmetric and ϕ is symmetric about the median of F, it is readily seen that eff(L)=0 . However, when F is not symmetric, but ϕ is symmetric about some value, (as is the case with most linear rank tests for scale), the integral in the numerator will be 0 only under very rare conditions. In practice, when commonly used score functions are used, the integral will be nonzero and the test will detect a difference in medians with skewed data when the null hypothesis of equal scale parameters is true. This is a major drawback of the commonly used rank tests for scale that is not shared by B .

4. Small Sample Comparisons.

This section examines the results of a Monte Carlo study used to compare the Type I error and

power properties of B_W, B_S, and the Siegel-Tukey test when sampling from a variety of population shapes. The distributions studied were standard normal, double exponential, uniform (-.5, .5), exponential, chi-square (d.f.=1), and a right-triangular distribution with density as shown in Table 1. It should be noted that the first three of these distributions are symmetric about 0 (zero) while the remaining three are shewed with left most end points at zero.

Throughout the study, observations for the two samples submitted to analysis, X' and Y', were obtained throught the transformations $X'=X$ and $Y'=cY+k\sigma_1$ where X and Y are randomly sampled from a common population with standard deviation σ_1. Five thousand repetitions of each experimental condition were employed. Simulation results are shown in Tables 3 through 11. Populations studied as well as values used for c and k are as indicated in these tables. Throughout the study each sample consisted of m=n observations. In evaluating the tabled results, it should be noted that $c=\sigma_2/\sigma_1=1$ (where σ_2 is the standard deviation of the Y') implies the null hypothesis while $c>1$ indicates the alternative. In addition, $k\neq 0$ implies a violation of assumption for S-T, but not for B_W or B_S.

The results for symmetric distributions are contained in Tables 3, 4, and 5 for the normal, the uniform, and the double exponential distributions, respectively. Table 3 shows that when sampling from the Gaussian distribution, all three tests produced Type I errors near the nominal level under the condition c=1 and k=0. However, as values of $k>0$ increase with c fixed at 1 (i.e. under a still true null hypothesis), S-T becomes increasingly conservative with the proportion of rejections reaching 0 (zero) for larger values of k. As expected, B_W and B_S continue to produce results near α. Under the condition k=0 and $c>1$, S-T is always more powerful than B_W and B_S. For $k\neq 0$, B_W and B_S are more competitive with B_S always being more powerful than S-T for $k\geq 2$. The magnitudes of the advantages of B_W and B_S over S-T are substantial in many cases with this being particularly true for B_S. It should also be noted that B_S is significantly more powerful than B_W in many circumstances.

Next, Table 4 shows that patterns of Type I error results obtained for the uniform distribution

were similar to those seen with the normal distribution. As might be expected from the ARE's reported in Table 2, the power advantages of S-T under this distribution are often somewhat larger than those obtained for the normal data. Once again, however, for k=3 large advantages are seen for B_W and B_S as compared to S-T with B_S being superior to B_W for most conditions. And from Table 5 it can be seen that the results for the double exponential distribution were similar to those found for the normal curve. It should be noted that for n=60 B_S was always more powerful than S-T for k>0.

Data for the skew distributions were generated by means of the same transformation model as was used for the symmetric data. For these distributions, however, the null condition was evaluated by fixing c at 1 while varying k from 0 to 3. This resulted in a simple shift of location. Tables 6 through 8 show similar patterns of Type I errors for the three statistics. Because of their insensitivity to location, B_W and B_S were valid for all conditions studied. By contrast, S-T was valid only for k=0. For k>0, S-T often produced highly inflated Type I error rates with inflation increasing with increases in sample size. It is also interesting to note that inflations were greatest for k=1 and diminished for larger k. The magnitudes of the inflations were also dependent upon the shape of the sampled population.

The alternative for skewed data was evaluated by setting k=0 and having c take the values 1.25, 1.50, 2.00 and 3.00. Since the means of all three skew distributions are greater than 0, these transformations generated data in which medians and variances depend on each other. Data with this characteristic are common in the biological sciences and are particularly common when populations are skewed. Tables 9 through 11 show that B_W and B_S were more powerful than S-T for all conditions studied. The magnitudes of these advantages were often quite large. As was true of the symmetric distributions, B_S was generally more powerful than B_W.

5. Discussion.

The distribution-free test for scale, B, is insensitive to, and therefore makes no assumptions

concerning, population locations. By contrast, the Siegel-Tukey test along with similar statistics is valid only in the highly restrictive and usually unrealistic circumstance that population medians are equal or are known. The tendency of S-T to become conservative with an accompanying loss of power when population medians differ is well known and often mentioned in the literature (see Boehnke (1989) for example). Researchers might be tempted, therefore, to apply the Siegel-Tukey test in situations where equivalence of population medians cannot be assumed with the expectation that, at worst, a conservative test will result. As has been demonstrated above, however, this test may become completely invalid when sampling is from skew distributions with Type I error rates approaching 1.0. This is particularly troublesome since it is precisely in these circumstances that a distribution-free test can be most useful.

We also reiterate that while the method of computing B_W and B_S effectively halves the sample sizes, these statistics remained competitive with S-T for $k>0$ and were often much more powerful than S-T in these situations. While the power advantages of B_S as compared to B_W were usually modest, they were nevertheless of sufficient size so as to favor this form of the statistic in applications. Use of B_s does not place an undo burden on the researcher since Savage scores are readily obtained through such popular statistics packages as the International Mathematical and Statistical Libraries (IMSL, 1987), SAS (SAS Institute, 1985) and SPSS (SPSS, 1990). A test statistic may be obtained by calculating an independent samples *t* test on the Savage scores with reference then being made to a *t* distribution with m+n-2 degrees of freedom. The *t* approximation for the sampling distribution of this statistic is quite good for the sample sizes employed in this study. (This was the manner in which B_S was used in this study.)

We recognize that the manner in which B_W and B_S are calculated does not provide a unique result. While this is not an entirely desirable trait, it should be noted that it is not unprecedented in rank-like statistics (cf. Moses (1963)), and that a "correct" result is obtained as long as the statistic is calculated in the manner described above. We believe that this negative trait is far outweighed the

fact that B is truly distribution-free in the sense that the Wilcoxon rank-sum test and other rank tests for location are distribution-free. That is, no distributional assumptions are made about symmetry or equal location parameters. This property is achieved by basing the test on the differences $X_i - X_j$ and $Y_i - Y_j$, as suggested by Theorem 2 in Bickel and Lehmann's (1979) discussion of measures of spread.

An interesting perspective on the test statistic B is obtained by noting the distinction between spread and dispersion made by Bickel and Lehmann (1979). Essentially, dispersion applies only to symmetric distributions and is relative to a location parameter, while spread is location free. In this respect, B is unique among rank-like tests for scale in that, being based on the differences $X_i - X_j$, it is a test for spread as opposed to being a test for dispersion.

6. Proofs.

Proof of Theorem 2.1. First, consider the case when σ approaches 0. The cdf of the smallest order statistic, $D_{x(1)}$, from the sample $D_{x1}, D_{x2}, \ldots, D_{x[m/2]}$, is $1-\left(1-F_{D_x}(t)\right)^{[m/2]}$ and the cdf of the largest order statistic, $D_{y([n/2])}$, from the sample $D_{y1}, D_{y2}, \ldots, D_{y[n/2]}$, is $\left(F_{D_y}(t)\right)^{[n/2]} = \left(F_{D_x}(t/\sigma)\right)^{[n/2]}$. Then, we have that

$$\Pr(D_{y([n/2])} < D_{x(1)}) = \int_0^\infty [m/2]\, f_{D_x}(t) \left(1-F_{D_x}(t)\right)^{[m/2]-1} \left(F_{D_y}(t)\right)^{[n/2]} dt$$
$$= E\left(\left(F_{D_y}(D_{x(1)})\right)^{[n/2]}\right) = E\left(\left(F_{D_x}(D_{x(1)}/\sigma)\right)^{[n/2]}\right).$$

Because $\lim_{t\to\infty} F_{D_x}(t) = 1$, it follows that $\lim_{\sigma\to 0} \Pr(D_{y([n/2])} < D_{x(1)}) = 1$. Hence, we have

$$\sup_{\theta\in\Omega\setminus\omega} \Pr(\text{reject } H_0 \mid \theta) \geq \lim_{\sigma\to 0} \Pr(\text{reject } H_0 \mid \sigma)$$
$$\geq \lim_{\sigma\to 0} \Pr(D_{y([n/2])} < D_{x(1)}) = 1,$$

which implies that B_N is resolving. By similar arguments, the theorem also holds when σ approaches infinity. Q. E. D.

References

Bickel, P. J. and Lehmann E. L. (1979), "Descriptive Statistics for Nonparametric Models IV. Spread," In: *Contributions to Statistics. Hajek Memorial Volume* (ed. by J. Jurekova). Academia, Prague, 33-40.

Boehnke, K. (1989), "A Nonparametric Test For Differences In The Dispersion of Dependent Samples," *Biometrical Journal*, 31, 421-430.

Bradley, J. V. (1968), *Distribution-Free Statistical Tests*, Englewood Cliffs, N.J: Prentice-Hall.

Duran, B. S. and Mielke, P. W. (1968), "Robustness of Sum of Squared Ranks Test," *Journal of the American Statistical Association*, 63, 338-344.

Fligner, M. A. and Killeen, T. J. (1976), "Distribution-Free Two-Sample Tests for Scale," *Journal of the American Statistical Association*, 71, 210-213.

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

International Mathematical and Statistical Libraries (1987), "*IMSL Library, Reference Manual*, Houston, TX: Author.

Jogdeo, K. (1966), "On Randomized Rank Score Procedures of Bell and Doksum," *Annals of Mathematical Statistics*, 37, 1697-1703.

Mielke, P. W. (1972), "Asymptotic Behavior of Two-Sample Tests Based on Powers of Ranks for Detecting Scale and Location Alternatives," *Journal of the American Statistical Association*, 67, 850-854.

Moses, L.E. (1963), "Rank Tests of Dispersion," *Annals of Mathematical Statistics*, 34, 973-983.

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

SAS Institute Inc. (1985), *SAS Users Guide* (5th ed.), Cary, NC: Author.

SPSS Inc. (1990), *SPSS Reference Guide*, Chicago, IL: Author.

Wasserstein, R. L. and Boyer, J. E. (1990), "Bounds on the Power of Linear Rank Tests for Scale Parameters," *The American Statistician*, 44, in press.

Table 1: Distribution of D_X

Distribution of X	Density of X	Density of D_X $(t>0)$
Uniform	$f(x)=1\ ;\ -\frac{1}{2}<x<\frac{1}{2}$	$f_{D_X}(t)=2-2t\ ;\ 0<t<1$
Normal	$f(x)=(2\pi)^{-1/2}\exp(-x^2/2)$	$f_{D_X}(t)=(\pi)^{-1/2}\exp(-t^2/4)$
Double Exponential	$f(x)=\exp(-\lvert x\rvert)/2$	$f_{D_X}(t)=\exp(-t)/2+t\ \exp(-t)/2$
Right Triangular	$f(x)=\frac{1}{2}x+1$	$f_{D_X}(t)=\frac{4}{3}-t+\frac{t^3}{12}$
Exponential	$f(x)=\exp(-x)\ ;\ x\geq 0$	$f_{D_X}(t)=\exp(-t)$

Table 2: Asymptotic Relative Efficiencies

Distribution	$ARE(B_N, S\text{-}T)$
Uniform	.222
Normal	.5
Double Exponential	.6328
Right Triangular	1.01
Exponential	3.35

Table 3: Comparisons of the Type I error and power properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from normal distributions and α=.05

			$\frac{\sigma_2}{\sigma_1}$				
n	k	statistics	1.00	1.25	1.50	2.00	3.00
20	0	B_W	.057	.079	.168	.346	.675
		B_S	.049	.078	.181	.395	.751
		S-T	.050	.110	.270	.588	.907
	1	B_W	.050	.089	.159	.359	.679
		B_S	.044	.089	.165	.407	.764
		S-T	.021	.054	.147	.458	.850
	2	B_W	.057	.085	.166	.348	.674
		B_S	.051	.081	.176	.394	.747
		S-T	.001	.003	.032	.208	.687
	3	B_W	.052	.079	.164	.357	.666
		B_S	.046	.077	.175	.402	.748
		S-T	.000	.000	.000	.037	.399
40	0	B_W	.051	.111	.281	.631	.936
		B_S	.054	.136	.349	.764	.983
		S-T	.048	.187	.473	.888	.998
	1	B_W	.046	.114	.280	.627	.932
		B_S	.042	.139	.361	.755	.981
		S-T	.019	.097	.332	.811	.995
	2	B_W	.050	.114	.276	.626	.933
		B_S	.050	.137	.352	.766	.982
		S-T	.001	.007	.071	.486	.959
	3	B_W	.049	.117	.283	.631	.927
		B_S	.048	.139	.361	.767	.980
		S-T	.000	.000	.001	.108	.780
60	0	B_W	.053	.147	.396	.798	.989
		B_S	.046	.197	.518	.923	.999
		S-T	.049	.263	.661	.979	1.000
	1	B_W	.051	.158	.404	.803	.989
		B_S	.052	.200	.519	.923	1.000
		S-T	.021	.147	.497	.943	1.000
	2	B_W	.049	.170	.389	.801	.989
		B_S	.046	.203	.519	.921	.999
		S-T	.000	.013	.133	.722	.997
	3	B_W	.050	.154	.386	.805	.989
		B_S	.044	.201	.518	.918	.999
		S-T	.000	.000	.003	.193	.937

Table 4: Comparisons of the Type I error and power properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from uniform distributions and α=.05

			$\frac{\sigma_2}{\sigma_1}$				
n	k	statistics	1.00	1.25	1.50	2.00	3.00
20	0	B_W	.057	.078	.171	.364	.680
		B_S	.047	.080	.191	.434	.776
		S-T	.049	.183	.424	.782	.966
	1	B_W	.055	.086	.154	.367	.678
		B_S	.050	.085	.182	.445	.781
		S-T	.017	.051	.194	.655	.941
	2	B_W	.054	.080	.173	.374	.679
		B_S	.050	.080	.199	.448	.778
		S-T	.000	.002	.012	.204	.846
	3	B_W	.050	.081	.165	.370	.666
		B_S	.045	.080	.187	.440	.776
		S-T	.000	.000	.000	.006	.530
40	0	B_W	.045	.120	.290	.648	.930
		B_S	.043	.164	.424	.819	.989
		S-T	.053	.341	.726	.978	1.000
	1	B_W	.048	.126	.286	.638	.930
		B_S	.046	.170	.421	.820	.989
		S-T	.012	.107	.429	.940	.999
	2	B_W	.049	.126	.290	.648	.928
		B_S	.043	.169	.429	.826	.986
		S-T	.000	.002	.031	.547	.994
	3	B_W	.047	.122	.291	.645	.936
		B_S	.049	.161	.424	.825	.989
		S-T	.000	.000	.000	.025	.905
60	0	B_W	.053	.162	.408	.824	.990
		B_S	.049	.248	.626	.954	1.000
		S-T	.050	.460	.888	.998	1.000
	1	B_W	.049	.161	.409	.825	.990
		B_S	.051	.257	.612	.959	.999
		S-T	.019	.156	.635	.992	1.000
	2	B_W	.045	.154	.401	.834	.988
		B_S	.042	.246	.620	.961	1.000
		S-T	.000	.003	.069	.792	1.000
	3	B_W	.050	.165	.411	.814	.989
		B_S	.051	.258	.636	.949	.999
		S-T	.000	.000	.000	.061	.988

Table 5: Comparisons of the Type I error and power properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from Laplace distributions and α=.05

			$\frac{\sigma_2}{\sigma_1}$				
n	k	statistics	1.00	1.25	1.50	2.00	3.00
20	0	B_W	.051	.082	.143	.300	.607
		B_S	.039	.075	.144	.315	.654
		S-T	.046	.082	.171	.404	.769
	1	B_W	.057	.072	.135	.303	.608
		B_S	.050	.067	.134	.322	.659
		S-T	.049	.064	.126	.340	.699
	2	B_W	.053	.081	.136	.297	.598
		B_S	.048	.074	.135	.319	.655
		S-T	.007	.016	.040	.161	.518
	3	B_W	.052	.074	.139	.288	.587
		B_S	.045	.073	.143	.310	.639
		S-T	.001	.001	.010	.052	.289
40	0	B_W	.048	.112	.225	.532	.882
		B_S	.047	.117	.265	.626	.947
		S-T	.051	.131	.328	.726	.980
	1	B_W	.047	.104	.219	.534	.888
		B_S	.047	.111	.260	.629	.951
		S-T	.046	.096	.246	.631	.956
	2	B_W	.048	.100	.217	.526	.893
		B_S	.044	.109	.255	.622	.955
		S-T	.008	.022	.081	.355	.839
	3	B_W	.054	.105	.219	.539	.893
		B_S	.052	.110	.260	.624	.947
		S-T	.000	.002	.013	.112	.610
60	0	B_W	.048	.137	.328	.708	.977
		B_S	.050	.153	.398	.820	.994
		S-T	.051	.182	.472	.888	.998
	1	B_W	.049	.130	.332	.719	.977
		B_S	.052	.149	.400	.824	.995
		S-T	.045	.128	.367	.812	.992
	2	B_W	.044	.136	.329	.716	.975
		B_S	.043	.155	.392	.825	.993
		S-T	.008	.035	.144	.554	.959
	3	B_W	.047	.128	.315	.722	.975
		B_S	.049	.146	.390	.817	.995
		S-T	.000	.002	.013	.186	.813

Table 6: Comparisons of the Type I error properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from an exponential distribution and α=.05

		$\frac{\mu_2-\mu_1}{\sigma}$			
n	statistics	0.0	1.0	2.0	3.0
20	B_W	.054	.054	.053	.050
	B_S	.050	.048	.044	.045
	S-T	.049	.449	.146	.017
40	B_W	.051	.050	.046	.047
	B_S	.052	.044	.044	.050
	S-T	.052	.816	.319	.034
60	B_W	.054	.045	.051	.048
	B_S	.047	.045	.052	.045
	S-T	.045	.961	.053	.041

Table 7: Comparisons of the Type I error properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from a chi-square (df=1) distribution and α=.05

		$\frac{\mu_2-\mu_1}{\sigma}$			
n	statistics	0.0	1.0	2.0	3.0
20	B_W	.048	.051	.050	.056
	B_S	.043	.045	.045	.048
	S-T	.051	.617	.190	.037
40	B_W	.048	.044	.048	.057
	B_S	.048	.045	.049	.060
	S-T	.054	.930	.373	.061
60	B_W	.050	.057	.053	.049
	B_S	.051	.057	.052	.048
	S-T	.050	.991	.559	.089

Table 8: Comparisons of the Type I error properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from a right- triangular distribution and α=.05

		$\frac{\mu_2-\mu_1}{\sigma}$			
n	statistics	0.0	1.0	2.0	3.0
20	B_W	.049	.051	.047	.056
	B_S	.044	.045	.037	.049
	S-T	.049	.087	.000	.000
40	B_W	.050	.051	.049	.048
	B_S	.050	.045	.047	.050
	S-T	.054	.214	.000	.000
60	B_W	.049	.051	.055	.055
	B_S	.045	.047	.052	.049
	S-T	.045	.367	.000	.000

Table 9: Comparisons of the power properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from an exponential distribution and α=.05

		$\frac{\sigma_2}{\sigma_1}$			
n	statistics	1.25	1.50	2.00	3.00
20	B_W	.070	.117	.239	.523
	B_S	.064	.115	.253	.557
	S-T	.060	.058	.066	.082
40	B_W	.086	.182	.436	.809
	B_S	.093	.214	.506	.890
	S-T	.053	.080	.114	.165
60	B_W	.112	.279	.623	.939
	B_S	.123	.326	.721	.977
	S-T	.070	.107	.175	.273

Table 10: Comparisons of the power properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from a chi-square (df=1) distribution and α=.05

		$\frac{\sigma_2}{\sigma_1}$			
n	statistics	1.25	1.50	2.00	3.00
20	B_W	.065	.100	.188	.361
	B_S	.063	.091	.194	.398
	S-T	.055	.058	.074	.104
40	B_W	.071	.136	.322	.648
	B_S	.078	.156	.377	.739
	S-T	.060	.080	.127	.206
60	B_W	.084	.184	.455	.817
	B_S	.100	.218	.539	.906
	S-T	.066	.090	.165	.322

Table 11: Comparisons of the power properties of the Siegel-Tukey, B_W and B_S tests for spread when sampling is from a right-triangular distribution and α=.05

		$\frac{\sigma_2}{\sigma_1}$			
n	statistics	1.25	1.50	2.00	3.00
20	B_W	.086	.156	.338	.640
	B_S	.083	.170	.399	.732
	S-T	.070	.100	.167	.189
40	B_W	.116	.249	.607	.919
	B_S	.146	.356	.766	.979
	S-T	.100	.207	.372	.486
60	B_W	.148	.380	.774	.986
	B_S	.204	.544	.925	1.000
	S-T	.137	.290	.582	.739