Precedence tests and confidence bounds for complete data : an overview and some results

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 94-19

Precedence Tests and Confidence Bounds for Complete Data: An Overview and Some Results

S. Chakraborti
P. van der Laan

Eindhoven, June 1994
The Netherlands

Precedence Tests and Confidence Bounds

for Complete Data:

An Overview and Some Results[1]

S. Chakraborti[1] and P. van der Laan[2]

[1] *University of Alabama, Department of Management Science and Statistics, P.O. Box 870226, Tuscaloosa, Alabama, U.S.A.*

[2] *Eindhoven University of Technology, Department of Mathematics and Computing Science, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

Summary

An overview of some nonparametric procedures based on precedence (or exceedance) statistics is given. The procedures include both tests and confidence intervals. In particular, the construction of some simple distribution-free confidence bounds for location difference of two distributions with the same shape is considered and some properties are derived. The asymptotic relative efficiency of an asymptotic form of the corresponding test relative to Wilcoxon's two-sample rank-sum test and the two-sample Student's t-test is given for various cases. Some K-sample problems are discussed where precedence type tests are useful, along with a review of the literature.

AMS Subject Classification: Primary 62G15, Secondary 62G25, 62G30.

Key words: Asymptotic Relative efficiency, Confidence bound, K-sample problem, Nonparametric, Precedence tests, Restricted alternatives, Two-sample problem.

[1]Support provided in part by NATO collaborative research grant CRG 920287.

1. Introduction

The main purpose of this paper is to present an overview of a class of simple nonparametric tests and confidence intervals based on what are called precedence (or exceedance) statistics.

A formal definition of a precedence statistic in the two-sample setting is as follows. Let there be two independent random samples of X's and Y's of sizes m and n, respectively, from two continuous populations. Let $X_{(r)}$ be a specified order statistic of the X-sample and let V_r denote the total number of Y-observations that do not exceed, that is precede, $X_{(r)}$. The statistic V_r is called a precedence statistic and a test based on V_r is referred to as a precedence test. A closely related test, used by some authors, is based on the statistic V_r', which denotes the number of Y's that exceed $X_{(r)}$. Thus V_r' is called an exceedance statistic and any test based on it is called an exceedance test. Clearly $V_r + V_r' = n$ so that the precedence and the exceedance tests are statistically equivalent. Interestingly, there is a connection between the count V_r and the rank R_r of $X_{(r)}$ in the combined sample of X's and Y's. This is given by $R_r = V_r + r$. Thus rank tests can be written in terms of precedence tests and vice-versa. Also, a precedence test can be viewed as a two-sample analog of the usual one-sample sign (quantile) test. For example, if we let $m \to \infty$, with $r/m \to p$ $(0 < p < 1)$, we arrive in the one-sample situation and the precedence statistic reduces to the well known quantile test statistic (the sign test obtains when $p = 1/2$). An indication of such a test appeared in Thompson (1938). A discussion on quantile tests can be found in standard books on nonparametric statistics (see for example Gibbons and Chakraborti, 1992) and as such the one-sample problem is omitted from this paper. Also, we concentrate on the K (≥ 2)-sample location problem and thus we do not consider applications of precedence or precedence type tests in the context of block designs.

We begin with a discussion of the two-sample location problem and review some of the literature on precedence (and related) tests. Some simple confidence bounds and the corresponding one-sided precedence tests based on pairs of order statistics are studied next. This is followed by a discussion on generalizations to the multi-sample location problem. An important practical advantage of the class of precedence tests, to be elaborated later, is that in certain experimental situations the tests can be applied before all the data are collected. Thus the experiment can be terminated early and a decision can be reached on the basis of a precedence test, resulting in savings in time and resources. Also, being distribution-free, the precedence tests and confidence bounds are more flexible and are more generally applicable. Moreover, the procedures share the advantage associated with many nonparametric procedures, namely that their applications only require simple calculations.

The paper is organized as follows. The two-sample problem is outlined in Section 2. Section 3 is a survey of some literature. The proposed confidence bounds are presented in Section 4 with a discussion on the selection of preferable bounds in Section 5. Some properties of the confidence bounds and the associated tests are examined in Section 6 and 7. In section 8 some K-sample problems are introduced. Various precedence type tests for these problems are reviewed in section 9.

2. The Two-sample Problem

Let X and Y be two independent random variables with unknown continuous cumulative distribution functions $F(x)$ and $G(y)$, respectively. The density functions corresponding to F and G, if they exist, are denoted by $f(x)$ and $g(y)$, respectively. Suppose two independent samples of observations $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$, respectively, are drawn from the two distributions so that there are a total of $N = m + n$ observations. The order statistics are denoted by $X_{(1)} < \cdots < X_{(m)}$ and $Y_{(1)} < \cdots < Y_{(n)}$, respectively. It is well known that the use of order statistics often results in very simple but efficient tests. Also, because of their very definition, the order statistics are often the most suitable statistics in a host of applications. For example, in the field of life-testing, when observations arrive in order of magnitude and one needs to analyze the data before all observations become available, it is natural to consider applying tests based on order statistics. The advantages are possible shortening of the duration of the experiment (testing time) and also of reducing the number of test items to be destroyed. We are interested in the location difference and it is assumed, without loss of generality, that $F(x) = G(x + \theta)$ for all x and with $\theta \in \Theta \subset \mathbb{R}^1$, so that $F(\cdot)$ and $G(\cdot)$ have the same shape. The problem is to consider distribution-free tests and the corresponding confidence bounds for θ based on some precedence statistics. The null hypothesis $H_0 : \theta = 0$, that the two distributions are identical, is tested against either the one-sided alternative $H_1 : \theta > 0$, or the two-sided alternative $H_2 : \theta \neq 0$.

It may be noted that a test based on the precedence statistic V_r is statistically equivalent to a test based on two order statistics, one from each sample. This follows since

$$V_r < i \quad \text{if and only if} \quad nG_n(X_{(r)}) < i \tag{1}$$

$$\text{if and only if} \quad X_{(r)} < Y_{(i)}, \tag{2}$$

where $G_n(.)$ is the empirical distribution function of the Y-sample. The last equation also highlights the fact that the precedence tests involve a comparison of the two-sample quantile functions. Various precedence tests have been proposed in the literature using either the counting form (1), or the order statistic form (2).

3. A Review of Some Literature

For two continuous populations with the same shape, precedence tests (and the corresponding confidence bounds) can be used to test for a possible difference in the location parameters. These tests are particularly useful when a simple and quick "in the field" analysis of the data is desirable. Some of the simplest tests will now be summarized.

For arbitrary sample sizes a simple procedure for testing is proposed by Tukey (1959). If one sample contains the highest value and the other the lowest, then we may choose i) to count the number of values in the one sample exceeding all values in the other, ii) to count the number of values in the other sample falling below all those in the one, and iii) to sum these two counts; it is required that neither count be zero. If m and n don't differ too much, then the critical values of the total count are roughly 7, 10 and 13 with two-sided confidence level of .05, .01 and .001, respectively. A one-sided test was also presented by Sidák-Vondrácek (1957).

A popular distribution-free test for the two-sample location problem is the Mood-Westenberg joint median test (Mood, 1954; Westenberg, 1948, 1950 and 1952). The test is based on W, the number of Y-observations smaller than (i.e., that precede) $\hat{m}$, the median of the combined sample of X's and Y's. The null hypothesis H_0 can be tested by comparing the proportions of the X's and the $Y's$ that lie below $\hat{m}$. If the alternative hypothesis is H_1, the test is to reject H_0 if $W < w$, where w is determined such that the size of the test is α. If the alternative is H_2, the test can be performed by setting up a 2×2-table with the X-sample and the Y-sample, as the row categories and "smaller than $\hat{m}$" and "larger than or equal to $\hat{m}$" as the column categories, and applying the usual chi-square criterion with 1 d.f. The Pitman

ARE (asymptotic relative efficiency) of the median test relative to Student's t-test in the case of two normal populations with equal variances is equal to $\frac{2}{\pi}$, which is about .64. The median test is the locally most powerful linear rank test when the underlying distributions are double exponential. Exact power computations for $m = n = 3, 4, 5$ and normal distributions have been made by Dixon (1954). Extensions to the case of several quantiles in the combined sample is considered by Massey (1951).

Chakravarti, Leone and Alanen (1962) have shown that the ARE of the median test and the test of Massey (1951) based on the first quartile and the median are zero, when these two tests are compared with the likelihood-ratio test for the exponential distribution. They found Massey's test to be about three times as efficient as the median test. Chakravarti, et al. (1961) derived the exact power of the median test and Massey's test under the exponential, the uniform shift and the uniform scale alternatives.

The median test is a special case of a joint quantile test. Let $Z_{(r)}$, where r is specified beforehand, be the r-th order statistic of the combined sample. The joint quantile test is based on the number of W_r, the number of Y-observations that are smaller than $Z_{(r)}$. Like the median test, the null hypothesis H_0 is rejected against the alternative H_1, if $W_r < w$, where w is obtained such that the size of the test is α. Against the two-sided alternative H_2, the test may be performed by setting up a 2×2-table with the X-sample and the Y-sample as the row categories and "smaller than $Z_{(r)}$" and "larger than or equal to $Z_{(r)}$" as the column categories and applying the usual chi-square criterion with 1 d.f. To see that a joint quantile test is a precedence test (Pratt, 1964; van der Laan, 1970b) note that $W_r < w$ if and only if the smallest $r - 1$ observations in the combined sample include at most $w - 1$ Y's and thus at least $r - w$ X's, which is possible if and only if $X_{(r-w)} < Y_{(w)}$.

Hemelrijk (1950) investigated joint quantile tests for arbitrary underlying distributions, continuous or discrete and proposed a generalization of the quantile test, namely a joint two-quantiles test. Here two percentiles $\hat{q}_{p_1}$ and $\hat{q}_{p_2} (> \hat{q}_{p_1})$ of the combined sample (for the exact definition, see Hemelrijk, 1950) are chosen and the number S_1 of observations of one sample smaller than $\hat{q}_{p_1}$ and the number S_2 of observations of the same sample larger than $\hat{q}_{p_2}$ are determined. The critical region of this test consists of pairs (S_1, S_2) with the smallest probabilities under H_0. Generalizations to more than two quantiles are possible.

Another simple test of the equality of two distributions is the control median test proposed by Mathisen (1943). The test is a precedence test based on V_{m+1}, the number of observations of the Y-sample of size $2n$ that precede the median of the X-sample of size $2m+1$. The null distribution of V_{m+1} can be obtained in a closed form and will be discussed later. Mathisen provided the exact lower and upper .01 and .05 percentage points for the test. For large m and n, a normal approximation can be used. The name control median test derives from the fact that in certain applications the X-population represents a control and the Y-population represents a treatment. The test is clearly applicable in situations where such a designation does not exist and in fact, in life-testing problems one can rename the samples and apply the test, depending on which of the sample medians become available first. Mathisen also discussed a precedence type test which makes use of the median and quartiles in the first sample.

It is clear that in general one may choose, instead of the median, say, the rth order statistic of the X-sample and base the test on the precedence statistic V_r. Such a test may be referred to as the control quantile test. Thus for the joint quantile test one finds a suitable quantile of the combined sample and uses the number of observations from any of the samples that precede the chosen quantile to define a test. On the other hand for the control quantile test one finds a suitable quantile from one of the samples (say the one corresponding to a control) and uses the number of observations in the other sample that precede the chosen control quantile to define a test.

Bowker (1944) showed that the control median test may not be consistent with respect to certain alternatives, say, when F and G are identical in the neighbourhood of their medians. However, the test can be shown to be consistent for the shift alternatives: $F(x) \equiv G(x+\theta)$, for the practically important case that $f(x) = F'(x)$ exists and the set $\{x; f(x) \neq 0\}$ is an interval. Similar remarks can be made for all these kinds of quantile tests.

Gart (1963) proposed an approximate chi-square test against a two-sided alternative based on

$$X^2 = \frac{(|2V_{s+1} - 1| \ - 1)^2(2s+3)}{n(2s+n+2)}, \tag{3}$$

where $m = 2s+1$. Under H_0, the distribution of X^2 can be approximated by a chi-square distribution with 1 d.f. Gart's test rejects H_0 in favour of H_2 if $X^2 > \chi^2_{1,\alpha}$. Gart also studied the asymptotic non-null distribution of the test statistic and showed that the Pitman ARE of the control median test is 1 with respect to the Mood-Westenberg joint median test. The same result is true for any other quantile. Gastwirth (1968) provided more details about the asymptotic distribution of the control median statistic and proposed a modification of the test better suited for the two-sided alternatives.

As noted earlier, precedence tests sometimes have the advantage that they lead to an early termination of an experiment. In the case of comparing life test results, where the data become available in a naturally increasing order of magnitude, it is possible that the unknown joint median is large, whereas the median of one of the samples, say of the X-sample, is not. In this situation the control median test has an important practical advantage over the joint median test, namely that the former can be applied before the latter (Gastwirth, 1968) and usually well before all the observations have been collected.

As noted before, in general, precedence tests can be based on the statistic V_r which equals the number of Y-observations that precede $X_{(r)}$, where $r = 1, 2, ..., m$, is selected in advance. It can be shown that under H_0

$$P(V_r = s) = \frac{\binom{r+s-1}{s}\binom{N-r-s}{n-s}}{\binom{N}{n}} \qquad s = 0, 1, ..., n, \tag{4}$$

and hence any precedence test based on V_r is a distribution-free test. Under the assumption $F(x) = G(x + \theta)$, X's are distributed as $Y - \theta$, so that under $H_1 : \theta > 0$, the X's are stochastically smaller than the Y's and therefore, intuitively, small values of V_r should mitigate against the null hypothesis in favour of the alternative. On the other hand if the alternative hypothesis is one-sided but in the opposite direction: $\theta < 0$, then the Y's are stochastically smaller than the X's and thus large values of V_r should lead to a rejection of H_0. These facts amply illustrate the simplicity and usefulness of precedence (count based) tests. A choice of the quantity r is important. A practical choice is the median of the X-sample, although ideally, r should be chosen so that the test is most "sensitive" to location differences for a variety of distributions and quantiles. This raises the question of whether there is a "best" precedence test; more will be said about this point later on. Some special cases of the precedence test are $r = \frac{1}{2}(m+1)$, m odd, which yields Mathisen's control median test, and $r = 1$, which yields the Rosenbaum (1954) test. Epstein (1954), Gumbel and Von Schelling (1950), Sarkadi (1957) and Harris (1952) presented derivations of null distributions and moments, as well as asymptotic approximations.

An expression for the power of a precedence test can be easily obtained. Suppose that the alternative is H_1, so that the test is to reject H_0 if $V_r < v$, where $v = v_\alpha$ is to be determined such that the size of the test is α. The power of the test follows from the fact that

$$P(V_r = i) = \frac{\binom{n}{i}}{B(r, m-r+1)} \int_0^1 (GF^{-1}(u))^i \{1 - GF^{-1}(u)\}^{n-i} u^{r-1} (1-u)^{m-r} du, \quad (5)$$

so that the power of the precedence test is given by

$$\beta(F, G) = \sum_{i=0}^{v_\alpha - 1} P(V_r = i), \quad (6)$$

where v_α is the largest integer such that

$$\sum_{s=0}^{v_\alpha - 1} P_{H_0}(V_r = s) \leq \alpha, \quad (7)$$

and $P_{H_0}(V_r = s)$, given by (4), follows from (5) when $F = G$.

Thus the power of a precedence test depends on the underlying distribution functions F and G only through the composite function $GF^{-1}(.)$. This shows that a precedence test is a strongly distribution-free test in the sense of Bell, Moser and Thompson (1966). Katzenbeisser (1989) and Liu (1992) presented results concerning the power of precedence tests against location alternatives under exponential, logistic and rectangular distributions. Sukhatme (1992) studied the power of some precedence tests under the Lehmann alternatives and obtained exact expressions for the powers of the Mathisen test and the Rosenbaum test. These results can be obtained from (6). For completeness it may be noted that closed form expressions for the power of precedence tests against the uniform shift alternatives, the exponential shift alternatives and the Lehmann alternatives can also be found in van der Laan (1970b). Also, it may be noted that the function GF^{-1} arises in the context of a two-sample P-P plot, a well known nonparametric graphical procedure (Wilk and Gnanadesikan, 1968) for testing if two distributions are identical. Young (1973) discussed the precedence test and obtained a normal approximation to its power function.

As noted earlier, the asymptotic distribution of the precedence statistic V has been obtained in the literature. Here we state the result as given in Chakraborti and Mukerjee (1989). Let ν_f be the pth quantile of the X-population and let $\{r_m\}$ be a sequence of positive integers such that $\lim_{m\to\infty}(r_m/m) \to p$, $0 < p < 1$. For example, one might take $r_m/m \le p < (r_m+1)/m$. Also let $\lim_{m,n\to\infty}(n/m) = \lambda$, where λ $(0 < \lambda < \infty)$ is a fixed quantity between 0 and ∞, and let $V_{m,n}$ be the number of Y-observations that do not exceed X_{r_m}. Then the random variable $n^{-1/2}[V_{m,n} - nG(\nu_f)]$ is asymptotically normally distributed with mean 0 and variance

$$G(\nu_f)\{1 - G(\nu_f)\} + \frac{g^2(\nu_f)}{f^2(\nu_f)}\lambda p(1-p). \tag{8}$$

A test statistic $(S_x + R_y) - (S_y + R_x)$ was suggested by Haga (1959/1960), where S_x and R_x denote the number of X-observations larger than $Y_{(n)}$ and smaller than $Y_{(1)}$, respectively. Similarly, S_y and R_y denote the number of Y-observations larger then $X_{(m)}$ and smaller than $X_{(1)}$, respectively. Hájèk and Sidák (1967) propose, among other forms of statistics, a test based on min $(S_x, R_y) - \min(S_y, R_x)$. These two test statistics may be used against one- and two-sided alternatives.

Epstein (1955) studied the relative merits of four nonparametric test procedures to test the null hypothesis of equal means, on the basis of samples of size 10 from two normal populations with equal variances. One of these tests is a special kind of an exceedance test for samples of equal size. Let $W_r = \max(X_{(r)}, Y_{(r)})$. If $W_r = X_{(r)}$, count the number of Y's which exceed $X_{(r)}$. On the other hand, if $W_r = Y_{(r)}$, count the number of X's which exceed $Y_{(r)}$. The test statistic E_r is the number of exceedances. The study was limited to the cases $r = 1$, 2 and 3. The other tests are the rank-sum test of Wilcoxon, the run test and the maximum-deviation test (this is a truncated maximum-deviation test (Tsao, 1954) with the truncation taking place at a time not later than $U_r = \max(X_{(r)}, Y_{(r)})$, where r is fixed in advance). In table 3.1 the experimental results for 200 pairs of samples are reproduced (the results for different rows are based on the same samples).

Insert Table 3.1 Here

Nelson (1963) proposed a precedence testing procedure particularly useful in life-testing problems. His test is based on the number K_1 of observations in the sample yielding the smallest observation which precede the r-th order statistic of the other sample. This test is mathematically equivalent to the exceedance test in which one counts the number K_2 of observations in the sample yielding the first failure which exceed the r-th order statistic of the other sample. The tests are related by $K_1 = n - K_2$ for all r, where n is the size of the sample yielding the smallest observation. Tables with critical values of K_1 for the precedence test with $r = 1$ are given for significance levels less than or equal to .10, .05, .01 (two-sided) and less than or equal to .05, .025, .005 (one-sided), for all combinations of sample sizes up to twenty. Recently, Nelson (1993) reintroduced the test and compared the power of a precedence test with the popular Wilcoxon rank-sum test in a simulation study where random samples were generated from two normal distributions. In this context it was reinforced that although a precedence test could lead to savings in resources by allowing a termination of the experiment before all the observations are made, the price one may have to pay is the power. For example, the power of a precedence test may be considerably less than that of the well known Wilcoxon rank-sum test, which of course, requires that all observations be available at the time of analysis. A computer program, written in BASIC, which can analyze the results of a precedence test, was provided.

Lehmann (1963) showed that if $P[U \leq a] = P[U \geq b] = \frac{\alpha}{2}$, then $P[D_{(a+1)} \leq \theta \leq D_{(b)}] = 1-\alpha$, where U is the Mann-Whitney statistic, $D_{(1)} < \ldots < D_{(mn)}$ are the ordered differences in observations from the two independent samples, and θ is the shift parameter. He further showed that this confidence interval, which, in a specific way, is related to a precedence statistic (van der Laan, 1970b; Chapter 7.1) inherits certain asymptotic efficiency properties from the Mann-Whitney test and gave a comparison of the interval $[D_{(a+1)}, D_{(b)}]$ with the t interval. Sen (1966) generalized this procedure from the Mann-Whitney to Chernoff-Savage type statistics. There are practical difficulties in using $[D_{(a+1)}, D_{(b)}]$ since many differences have to be ordered. This can be partially alleviated by some graphical methods presented in Moses (1965) and discussed in several textbooks (e.g. Gibbons, 1976; Gibbons and Chakraborti, 1992).

Eilbott and Nadler (1965) investigated some precedence tests for life-testing problems under the assumption of underlying exponential distributions. Let $F(x) = 1 - \exp(-x/\theta_x), x > 0$, and $G(y) = 1 - \exp(-y/\theta_y)$, $y > 0$, respectively, and suppose $H_0 : \theta_x = \theta_y$ is to be tested against $H_1 : \theta_x > \theta_y$. The two groups of items (of size m and n, respectively) are placed on test simultaneously and testing is terminated as soon as k out of the m X's fail or r out of the n Y's fail, whichever comes first. The null hypothesis is rejected in favour of the alternative if k of the X failures are observed before r of the Y failures. Note that this is a precedence test since the critical region is equivalent to $V_r^* \geq k$, where V_r^* denotes the number of X's that precede $Y_{(r)}$. They discussed various properties of the precedence test including the expected duration (testing time) and obtained closed-form expressions for the power function. They also derived the one-sided uniformly most powerful test for this problem and compared it, asymptotically, with the precedence test. For the problem of testing H_0 against $H_2 : \theta_x \neq \theta_y$, Eilbott and Nadler proposed to reject the null hypothesis if $X_{(k_1)} < Y_{(r_1)}$ or $Y_{(k_2)} < X_{(r_2)}$. If, in addition, the restriction: $\min(k_1, k_2) \geq \max(r_1, r_2)$ is imposed, then the power function of the test is given by $P(X_{(k_1)} < Y_{(r_1)}) + P(Y_{(k_2)} < X_{(r_2)})$. For the special case $r_1 = r_2 = r$, $k_1 = k_2 = k$ and $m = n$, these restricted test plans are equivalent to the procedures investigated by Epstein (1955). On the other hand, when $r_1 = r_2 = 1$ does *not* hold, these restricted test plans differ from the general two-tailed tests proposed by Nelson (1963). Nelson's procedure depends on which sample gave rise to the first observed failure, whereas Eilbott and Nadler's procedure clearly does not. Their findings provide further insight into its properties in situations where the underlying distributions are unknown. They concluded that "...unless one knows that the underlying distributions are approximately normal, the use of a precedence life test with very small r is unwise. If no pertinent knowledge of the underlying distributions is available, we believe it prudent to choose a test plan with $r \geq 3$ and balanced sample sizes when circumstances permit."

Shorack (1967) showed that the expressions of the power function derived by Eilbott and Nadler are in fact valid for a large class of distributions which include the exponential distribution, namely the class of distributions $\mathcal{F} = \{(F, G) : G = 1 - (1 - F)^\delta, \delta > 0\}$. He showed that the power function in the case of exponential distributions with difference in scale parameters is a function of $\lambda = \theta_y / \theta_x$ only. Young (1973) presented some asymptotic results for the precedence test of Eilbott and Nadler (1965), including the asymptotic relative efficiency with respect to the F test for the scale parameter of an exponential distribution.

The precedence statistic V_r, which represents the number of Y's that precede $X_{(r)}$, is also referred to as the "placement" of $X_{(r)}$ among the observations in the Y-sample. Fligner and Wolfe (1976) and Orban and Wolfe (1982) derived various properties of the placements, including the first two moments under the null hypothesis. The placements are closely related to another class of statistics called the "block-frequencies" or the "two-sample coverages." The reader is referred to Wilks (1963) for a discussion of the statistical properties of the coverages.

Hackl and Katzenbeisser (1984) proposed a precedence type test of H_0 against the alternative that the dispersion of F exceeds the dispersion of G. Chakraborti and Mukerjee (1989) considered the problem of estimating the probability that a Y-observation will exceed some specified quantile of the X-population. Such a quantity can be used to define a nonparametric measure of the difference between F and G and may be useful in survival/reliability analysis, especially in situations where the X-population represents a "control" and the Y-population represents some experimental condition. An asymptotically distribution-free confidence interval for this measure was given based on a precedence type statistic and the performance of the interval was studied in a simulation study.

Lin and Sukhatme (1992) considered the problem of finding the "best" precedence test for the "Lehmann alternatives" where distribution functions F and G are related by $F(x) = 1 - (1 - G(x))^{\lambda}, \quad \lambda > 1$. The Lehmann alternatives, which define a class of distributions, allow a semiparametric formulation of the equality of two distributions and are quite popular in survival analysis and reliability problems. The class includes exponential distributions, Weibull distributions differing only in scale and distributions with proportional hazard rates. The "best" means the most powerful test against a simple alternative $\lambda = \lambda_0$. This approach is different from the one in van der Laan (1970b) where the idea of a most stringent (Lehmann, 1959) test is used.

Finally, it may be noted that some authors have studied precedence tests for the Behrens-Fisher type problems. For example, Schlittgen (1979) proposed a nonparametric test for testing differences of location in two independent samples without assuming equal scale parameters. The test is based on simultaneous application of two modified median tests, each using its sample median of the two samples. The test is somewhat complicated since it is based on a two-dimensional rejection region. In this paper, however, we concentrate on the shift alternatives.

4. Simple Confidence Bounds and Tests Based on Pairs of Order Statistics

In this section confidence bounds and corresponding tests based on pairs of order statistics will be considered for the two-sample case. The corresponding tests are precedence tests. The presentation is from van der Laan (1970a,b) where a general class of lower confidence bounds was considered for the difference in location, θ, defined by

$$P(\beta\text{th largest of } Y_{(j_r)} - X_{(i_r)} < \theta;\ r = 1, \cdots, R) \geq 1 - \alpha_l, \tag{9}$$

and a class of upper confidence bounds for θ defined by

$$P(\gamma\text{th largest of } Y_{(j_r)} - X_{(i_r)} > \theta;\ r = 1, \cdots, R) \geq 1 - \alpha_u, \tag{10}$$

with $1 \leq R \leq mn,\ 1 \leq \beta,\ \gamma \leq R,\ 1 \leq i_1, \cdots, i_R \leq m$ and $1 \leq j_1, \cdots, j_R \leq n$.

If $R = 1$, and thus $\beta = 1$, then a lower confidence bound for θ is given by $Y_{(j_1)} - X_{(i_1)} < \theta$, where i_1 and j_1, can be chosen in such a way that the confidence bound and/or the corresponding test have nice properties. Similar remarks can be made for upper confidence bounds for θ.

If the lower confidence bound is smaller than the upper confidence bound, with probability one, then a confidence interval with confidence level α can be obtained by combining these bounds, and where $\alpha = \alpha_l + \alpha_u$. For the confidence coefficients $1 - \alpha_l^*$, $1 - \alpha_u^*$ and $1 - \alpha^*$ the property $\alpha^* = \alpha_l^* + \alpha_u^*$ holds.

The case $R = 1$ is now considered in more detail. In this case the class of lower confidence bounds is denoted by C_1 and the class of corresponding precedence tests is denoted by V_1. The mn lower confidence bounds $Y_{(j)} - X_{(i)}$ of C_1 for θ are denoted by D_{ji}, with $1 \leq i \leq m$ and $1 \leq j \leq n$, and are given in Mood (1950) and Mood and Graybill (1963). Mood and Graybill suggest a criterion for selecting a particular confidence bound to be used. This criterion is such that the corresponding test has significance level approximately equal to α.

The confidence coefficient of the lower confidence bound D_{ji} is equal to

$$P(D_{ji} < \theta) = \frac{1}{\binom{N}{n}} \sum_{l=j}^{n} \binom{i+l-1}{l} \binom{N-i-l}{n-l}, \quad i = 1, ..., m;\ j = 1, ..., n, \tag{11}$$

which follows from (4) using the fact that

$$P(D_{ji} < \theta) = P(Y_{(j)} - X_{(i)} < \theta) = P(Y'_{(j)} < X_{(i)}),$$

with $Y'_{(j)} = Y_{(j)} - \theta$, $j = 1, ..., n$. The lower confidence bound yields a test of $H_0 : \theta = 0$ against the alternative $H_1 : \theta > 0$. The test rejects H_0 if $D_{ji} > 0$. Note that $D_{ji} > 0$ if and only if $Y_{(j)} > X_{(i)}$, so that the D_{ji} test is a precedence test. Without loss of generality the null hypothesis can be formulated as $\theta = 0$, because $\theta = \theta_0$ (θ_0 known) can be transformed into H_0 by taking $Y - \theta_0$ instead of Y.

The following theorem shows that the precedence tests are consistent.

Theorem 4.1. If m, n, i and j tend to infinity such that $\frac{i}{m}$, $\frac{j}{n} \to \lambda$ $(0 < \lambda < 1)$, and $F(\cdot)$ has a density $f(\cdot)$ continuous in the neighbourhood of ξ_λ, where $F(\xi_\lambda) = \lambda$ and $f(\xi_\lambda) > 0$, then $P(D_{ji} > 0 \mid H_1 : \theta = \theta_1 > 0)$ tends to 1.

Proof:

Under the given conditions $Y_{(j)}$ and $X_{(i)}$ converge in probability to η_λ and ξ_λ, respectively, where $F(\xi_\lambda) = \lambda$ and $G(\eta_\lambda) = \lambda$. So for each $\varepsilon > 0$ and $0 < \delta < 1$ one can find $N_{\varepsilon,\delta}$ such that for m and n larger than $N_{\varepsilon,\delta}$ one has

$$P\left(|Y_{(j)} - \eta_\lambda| < \frac{\varepsilon}{3}\right) > 1 - \delta$$

and

$$P\left(|X_{(i)} - \xi_\lambda| < \frac{\varepsilon}{3}\right) > 1 - \delta\ .$$

Now under $H_1, \eta_\lambda > \xi_\lambda$, so that taking $\varepsilon = \eta_\lambda - \xi_\lambda = \theta_1$, one gets

$$\begin{aligned} P[D_{ji} > 0] &= P[Y_{(j)} - X_{(i)} > 0] \\ &\geq P\left[|Y_{(j)} - \eta_\lambda| < \tfrac{\varepsilon}{3} \wedge |X_{(i)} - \xi_\lambda| < \tfrac{\varepsilon}{3}\right] \\ &= P\left[|Y_{(j)} - \eta_\lambda| < \tfrac{\varepsilon}{3}\right] P\left[|X_{(i)} - \xi_\lambda| < \tfrac{\varepsilon}{3}\right] > (1-\delta)^2\ , \end{aligned}$$

and the result follows immediately. □

The subscripts j and i of D_{ji} must be chosen beforehand, namely before the results of the experiment are known. In the next section a selection scheme for the subscripts is indicated.

5. Selection of Confidence Bounds from C_1

In C_1 there are many candidates for a lower confidence bound for θ. It is possible to select a lower confidence bound on the basis of the idea of a most stringent test.

For any pair of sample sizes (m, n) and a given level of significance α, the test with the maximal size among all available level α tests D_{ji} is determined[2]. Denote the power function of this test by $\beta_{\text{max size}}(\theta; \alpha)$. Then $\theta'_{.50}$ is determined for which $\beta_{\text{max size}}(\theta'_{.50}, ; \alpha) = .50$. Next among all level α tests D_{ji} the one with maximal power at $\theta'_{.50}$ is determined and its power function is denoted by $\beta^*(\theta)$. Then $\theta^*_{.25}, \theta^*_{.50}$ and $\theta^*_{.75}$ are calculated by $\beta^*(\theta^*_{.25}) = .25$, $\beta^*(\theta^*_{.50}) = .50$ and $\beta^*(\theta^*_{.75}) = .75$. These computations are carried out in order to find the region in which the power functions of the better tests have interesting values. For all level α tests the power is determined at these three points $\theta^*_{.25}$, $\theta^*_{.50}$ and $\theta^*_{.75}$. In each of these three points the maximal powers $b_{.25}$, $b_{.50}$ and $b_{.75}$, respectively, among all level α tests are determined. Then the test being selected is the one (with power function $\beta(\theta)$) among all level α tests for which

$$(b_{.25} - \beta(\theta^*_{.25})) + (b_{.50} - \beta(\theta^*_{.50})) + (b_{.75} - \beta(\theta^*_{.75}))$$

is minimal, i.e. minimizes the average shortcoming over three interesting points. Thus, roughly speaking, this selected test has, on average, optimal power among all level α tests.

This selection of tests, and consequently of lower confidence bounds, has been performed for normal distributions with common a variance σ^2 and for sample sizes (≥ 3) up to and including $m = 15$ and $n = 15$ and for six significance levels. In this paper only the results for two significance levels, .01 and .05, are given in Tables 5.3 and 5.4. We need to compute the power $\beta(\delta), \delta = \theta^*/\sigma$, of the D_{ji} test for testing H_0 against $H_1 : \theta = \theta^* > 0$. From (6) one gets

$$\beta(\delta) = \sum_{i=0}^{v_\alpha - 1} \frac{\binom{n}{i}}{B(r, m-r+1)} \int_{-\infty}^{\infty} \Phi^i(x-\delta)\Phi^{n-i}(\delta - x)\Phi^{r-1}(x)\Phi^{m-r}(-x)d\Phi(x), \quad (12)$$

[2]In some cases there is more than one test with maximal size. In these cases the test with maximal i has been taken.

where Φ denotes the standard normal distribution function. To evaluate (12) numerical integration has been employed where an approximation for $\Phi(\cdot)$ (cf. Hastings, 1955) has been used. The error function has been approximated with maximal error 10^{-9}. Note that if D_{ji} is the selected test for the pair of sample sizes m and n, then $D_{m-i+1,n-j+1}$ is the selected test for the pair of sample sizes n and m. For large m and n normal approximations can be used (cf. Mood, 1950; Mood and Graybill, 1963; van der Laan, 1970b).
Using standard methods it can be shown that in general the ARE of the D_{ji} test relative to the Wilcoxon rank-sum test is equal to

$$\frac{f^2(0)}{3(\int_{-\infty}^{\infty} f^2(t)dt)^2}, \tag{13}$$

for any continuous density function f with $f(0) > 0$. Relative to the two-sample t-test the ARE of the D_{ji} test is equal to

$$4\sigma^2 f^2(0), \tag{14}$$

for any continuous density function f with variance σ^2.

In Tables 5.3 and 5.4 some values of the ARE are given.

Insert Table 5.3 and Table 5.4 Here

6. Comparison of the Selected Lower Confidence Bounds D_{ji} with the Lower Confidence Intervals Based on the t-distribution

It is possible to compare the selected lower confidence bounds with the lower confidence bounds based on the t-distribution directly by means of the expected lengths of the one-sided confidence intervals. The length of a one-sided confidence interval is defined in this context as the absolute value of the difference of θ and the confidence bound. In the case of normal distributions with common variance 1 (this assumption can be made without loss of generality) and confidence level $1 - \alpha = .95$ the selected lower confidence bounds D_{ji} (with confidence coefficient $1 - \alpha^*$) have been compared with the lower confidence bounds based on the two-sample Student's t-test with the same confidence coefficient for each pair of sample sizes (smaller than or equal to 15) by determining the ratios of the expected lengths of the one-sided confidence intervals, namely

$$E\{\theta - (Y_{(j)} - X_{(i)})\} \times (E\left\{\theta - \left[\bar{Y} - \bar{X} - t_{1-\alpha^*}(N-2)\left(\frac{N}{mn}S^2\right)^{\frac{1}{2}}\right]\right\})^{-1}, \tag{15}$$

with S^2 is the pooled estimator of the common variance and where the $(1-\alpha^*)$-percentage points $t_{1-\alpha^*}(N-2)$ of the t-distribution with $(N-2)$ degrees of freedom and $E\{S\}$ have been approximated by some simple extra- and interpolations from tables in Pearson and Hartley (1958). The results are presented in table 6.1. It can be seen that for this case the loss is about 20 per cent, which seems satisfactory.

Insert Table 6.1 Here

Next we discuss some multi-sample extensions of precedence tests.

7. Some K-sample Problems

Suppose that K (≥ 2) independent random samples X_{hi}, $i = 1, ..., n_h$ and $h = 1, ..., K$ are available from continuous distributions with distribution functions $F_1, ..., F_K$, respectively, and the null hypothesis of homogeneity

$$H_0 : F_1 = F_2 = ... = F_K$$

is to be tested against some alternative hypothesis. For the usual K-sample problem, the alternative hypothesis is often the global alternative

$$H_1 : \textit{not } H_0$$

and the most popular nonparametric test is the Kruskal-Wallis test (Lehmann, 1975). In some situations however, the alternative hypothesis specifies some order relationship among the distributions and with this information available a-priori, one should be able to design specific tests that are more powerful than the Kruskal-Wallis test. For example, if F_1 corresponds to some "control", one may be interested in testing H_0 against the partially-ordered alternative

$$H_2 : F_i \leq F_1, i = 2, ..., K ,$$

with strict inequality for some i. This would be the case when one wishes to test if any of the "treatments" $2, ..., K$, is "better" than the control. In the literature of "order restricted statistical inference" (Robertson, Wright and Dykstra, 1989), this is referred to as the simple-tree alternatives or the many-to-one problem.

On the other hand, if the K treatments respresent, for example, the increasing dose levels $(1, ..., K)$ of some drug, one may be interested to test if the responses under these treatments also exhibit an increasing order. This is called the upward trend problem or the problem of simple order. The alternative hypothesis in this case is written as

$$H_3 : F_1 \geq F_2 \geq ... \geq F_K \ ,$$

with at least one strict inequality. In the next section we discuss some precedence type tests for these problems which may be viewed as generalizations of their two-sample counterparts.

8. A Review of Some Literature

First consider the problem of testing H_0 against the global alternative H_1. Massey (1951) considered a multi-sample extension of the median test for this problem using several order statistics of the combined sample. Suppose that there are no ties and of the $N = \sum_{h=1}^{K} n_h$ ordered observations in the combined sample, $r-1$ are chosen. These are denoted by $Z_{\alpha_1}, ..., Z_{\alpha_{r-1}}$, where the α_i are integers with $1 \leq \alpha_1 < ... < \alpha_{r-1} \leq N$. Let W_{hj} denote the number of observations such that $Z_{\alpha_{j-1}} < X_{hi} \leq Z_{\alpha_j}$, $j = 2, ..., r-1$, $h = 1, ..., K$. Also let W_{h0} denote the number of $X_{hi} \leq Z_{\alpha_1}$ and let W_{hr} denote the number of $X_{hi} > Z_{\alpha_{r-1}}$. Under the null hypothesis the joint distribution of the W's is given by Massey (1951) which does not depend on the distribution F_i. Therefore tests based on the W's are distribution-free.

A direct extension of the Mood-Westenberg joint median test is obtained by choosing only the median in the combined sample and noting the number of observations $W_1, ..., W_K$, respectively, from each of the K samples, that precede the median of the combined sample. Under the null hypothesis the joint distribution of the W's can be shown to be

$$f(w_1, ..., w_K) = \frac{\binom{n_1}{w_1} ... \binom{n_K}{w_K}}{\binom{N}{t}}, \qquad (16)$$

where $t = N/2$ if N is even and $t = (N-1)/2$ if N is odd. The Mood-Brown-Westenberg median test of H_0 against H_1 is based on

$$Q_N = \frac{N(N-1)}{t(N-t)} \sum_{i=1}^{K} \frac{(w_i - \frac{n_i t}{N})^2}{n_i}, \tag{17}$$

and the approximately size α test rejects H_0 in favour of H_1 if

$$Q_N \geq \chi^2_{K-1,\alpha}. \tag{18}$$

Andrews (1954) derived the efficacy of the Mood-Brown-Westenberg median test and showed that with respect to the Kruskal-Wallis (global) test the ARE is equal to the ARE of the sign test relative to the Wilcoxon signed rank test, given in (13). Thus, for example, the ARE is 2/3 when the underlying distributions are normal.

An extension of the control median test for the multi-sample problem is as follows. Let M be the median of the control sample (say the first sample) and let $V_{i0}, i = 1, ..., K$, denote the number of observations under the ith treatment that precede M. Under the null hypothesis, the joint distribution of the V_{i0}'s can be shown (see Sen, 1962; also Gibbons and Chakraborti, 1992; chapter 11) to be

$$f(v_{20}, ..., v_{K0}) = \frac{\prod_{i=1}^{K} n_i!}{N!} \frac{v_1!}{\prod_{i=1}^{K} v_{i1}!} \frac{(v_0 - 1)!}{(v_{10}-1)! \prod_{i=2}^{K} v_{i0}!}, \tag{19}$$

where $v_j = \sum_{i=1}^{K} v_{ij}, j = 0, 1$, and $v_{i1} = n_i - v_{i0}$. Thus the null distribution does not depend on the underlying distribution and so any test based on the V_{i0}'s is distribution-free. The asymptotic joint distribution of the V_{i0}'s can be shown to be a multivariate normal distribution. As with the median test, a large sample test of H_0 against H_1 can be based on a quadratic form in the V_{i0}'s with the large sample covariance matrix as the discriminant, the rejection region of the test consisting of large values of the test statistic. Sen (1962) proposed the criterion

$$C_N = \frac{v_{10}}{(n_1+1)} \sum_{i=1}^{K} n_i \left(\frac{v_{i0}}{n_i} - \frac{v_0}{N} \right)^2. \tag{20}$$

Under H_0 and some mild conditions, asymptotically, C_N has a chi-square distribution with $K-1$ degrees of freedom. An approximately size α test of H_0 against H_1 is

$$C_N > \chi^2_{K-1,\alpha}. \tag{21}$$

Sen (1962) showed that the ARE of the C_N test relative to the Q_N test is 1, so that for large sample sizes the two median tests are 'power equivalent.' Extensions to the cases where some quantile other than the median is of interest or where one picks two or more quantiles in the control sample instead of just one are clearly plausible and have been studied in the literature. Among these, Sen (1962) treated the general case, which is also discussed in Gibbons and Chakraborti (1992; Chapter 11), where additional references can be found. Sen (1962) provided asymptotic power and efficacy calculations together with suggestions about a choice of the number and the order of the control quantiles to define the "best" test. The consensus is that one or two quantiles is usually sufficient in practice.

Slivka (1970) considered a test for the simple-tree alternatives problem based on the counts $V_{i0}, i = 2, ..., K$. Let $V = \min(V_{20}, ..., V_{K0})$. Slivka's test uses the union-intersection principle and rejects H_0 in favour of H_2 if at least one of the V_{i0}'s is sufficiently small, that is if V is small. The null distribution of V is given by

$$P(V \leq v|H_0) = \sum_{h=2}^{K}(-1)^{h-2}\begin{pmatrix} K \\ h \end{pmatrix}\sum_{v_{20}=0}^{v} \cdots \sum_{v_{h0}=0}^{v} f(v_{20}, ..., v_{h0}|H_0). \tag{22}$$

Slivka provided tables for the exact P-values when $n_1 = 1(1)14$ and $n_2 = ... = n_K = n = [1/3n_1](1)n_1$, (where $[a]$ represents the largest integer not exceeding a) and $K = 2(1)9$. Slivka also considered the asymptotic null distribution of V and obtained a normal approximation. When $n_i = \lambda n_1$, where $i = 2, ..., K$ and λ is some constant, the probability $P(V \leq v|H_0)$ can be approximated by $1 - G(H; N, \rho)$, where $G(H; N, \rho)$ is the quantity tabulated by Gupta (1963) (in the context of the cumulative distribution function of the maximum of N normal random variables with mean 0, variance 1, and a common correlation coefficient). In our case the quantities H, N and ρ are,

$$-\frac{\{(n_1+1)V - ns\}}{ns(n_1 - s + 1)(n_1 + n + 1)/(n_1 + 2)}, \tag{23}$$

$K - 1$ and $n/(n_1 + n + 1)$, respectively. When only critical values for the test are desired, one can use the tables given, for example, in Bechhofer and Dunnett (1988).

Chakraborti and Desu (1988a) considered an extension of the control quantile test for the simple-tree problem. Their test is based on $T = \sum_{i=2}^{K} V_{i0}$ and the test rejects H_0 in favour of H_2 if T is small. Since T counts the number of treatment observations, out of a total of $N^* = N - n_1$, that precede the rth control order statistic, the null distribution of T can be obtained from the null distribution of a two-sample precedence statistic, given in (4), with $n = N^*$ and $m = n_1$. For small sample sizes, Kao and Chakraborti (1994) developed a computer program to tabulate the exact critical values. A normal approximation, given in Chakraborti and Desu (1988a), is adequate for moderate to large sample sizes. Simulation studies have indicated that the test of Chakrabori and Desu is more powerful than Slivka's test when the treatments are about equally effective but more effective than the control, whereas Slivka's test is more powerful in situations where only some of the treatments are better than the control.

It may be noted that in practice, situations arise where one would like to test whether the quantiles of K treatment populations are equal to some specified (standard) value against the alternative that at least one of the treatment quantiles is greater (or smaller). In this case one can use precedence type statistics, which in fact, are the familiar sign test statistics. We will not discuss any of these details; readers are referred to Chakraborti and Gibbons (1991, 1992, 1993), Chakraborti (1991) and Ismail (1992).

Next we consider the problem of simple-order or the upward trend problem. Gore, Rao and Sahasrabudhe (1986) proposed an extension of the joint median test based on $T^* = \sum_{i=1}^{K} \frac{iW_i}{n_i}$, where recall that W_i denotes the number of observations from the ith sample that precede the median of the combined sample. The test rejects the null hypothesis if T^* is large. The asymptotic distribution of T^* can be shown to be normal. In particular, under the null hypothesis, the asymptotic mean and the variance of T^* equals $tK(K+1)/2N$ and

$$\{\sum_{i=1}^{K}(i/n_i)^2\tau_i - (\sum_{i=1}^{K} i\tau_i/n_i)^2\}/4, \tag{24}$$

respectively, where $\tau_i = \lim_{N\to\infty}(n_i/N)$. These moments can be used to define an approximately size α test. The authors also considered the asymptotic nonnull distribution of T^* under a sequence of local translation alternatives and derived the Pitman ARE against the well known Jonckheere-Terpstra test. It is interesting to note that when the location parameters are equally spaced, the ARE turns out to be the same as that of the Mood-Brown-Westenberg joint median test against the Kruskal-Wallis test, given in (13). One can further consider some quantile other than the median of the combined sample and construct a linear function of the counts. Distributional properties of such a 'linear joint quantile' test can be obtained along similar lines.

In some situations a practitioner may like to use a linear function of the W's, say, $\sum_{i=1}^{K} a_i W_i$ as a test statistic, where the weights a_i are determined such that the resulting test has some optimal (power) property. To this end one approach used in the literature is to maximize the Pitman efficacy (asymptotic local power) of the class of tests against a sequence of local alternatives (converging to the null hypothesis). It has been shown (Rao and Gore, 1984) that in the balanced (equal sample size) design case, the optimal weight a_i is proportional to i, $i = 1, ..., K$, when the location parameters are equally spaced. One can also consider deriving weights that optimize other aspects related to the power of the test. For example, one could attempt to find the weights that minimizes the maximum shortcoming (Schaafsma, 1966) of the test. It will be interesting to compare these tests, especially for small to moderate sample sizes.

Chakraborti and Desu (1988b) considered a linear function of the precedence statistics, $\sum_{i=2}^{K} b_i V_{i0}$, for the simple-order problem and obtained the optimal weights which maximize the Pitman efficacy against a class of local translation alternatives. The results are similar to those for the median statistics, namely that for equal sample sizes and equal spacings the optimal weight b_i is proportional to $i = 2, ..., K$.. When the median is the quantile of interest, with equal sample sizes and equal spacings, the ARE of the optimal member of the Chakraborti-Desu class of statistics relative to the popular tests of Jonckheere-Terpstra (and Rao-Gore) can be shown to be equal to the ARE of the sign test relative to the Wilcoxon signed rank test, given in (13). Thus, for heavy-tailed distributions like the Cauchy and the double exponential, the optimal member of the Chakraborti-Desu class of tests is more efficient.

References

Andrews, F. C. (1954). Asymptotic behavior of some rank tests for analysis of variance. Annals of Mathematical Statistics **25**, 724-736.

Bechhofer, R.E. and C.W. Dunnett (1988). Tables of percentage points of the multivariate Student t distribution. Selected Tables in Mathematical Statistics **11**, Institute of

Mathematical Statistics, Providence, Rhode Island.

Bell, C.B., J.M. Moser and R. Thompson (1966). Goodness criteria for two-sample distribution-free tests. Annals of Mathematical Statistics **37**, 133-142.

Bowker, A.H. (1944). Note on consistency of a proposed test for the problem of two samples. Annals of Mathematical Statistics **15**, 98-101.

Chakraborti, S. (1991). Graphical comparison of means with a standard in the one-way layout. Journal of Applied Statistics **45**, 227-254.

Chakraborti, S. and M.M. Desu (1988a). Generalizations of Mathisen's median test for comparing several treatments with a control. Communications in Statistics: Simulation and Computation **17**, 947-967.

Chakraborti, S. and M.M. Desu (1988b). A class of distribution-free tests for testing homogeneity against ordered alternatives. Statistics and Probability Letters **6**, 251-256.

Chakraborti, S. and J.D. Gibbons (1993). Nonparametric comparison of treatments with a standard in the one-way layout: some design considerations. Communications in Statistics: Theory and Methods **22**, 1-14.

Chakraborti, S. and J.D. Gibbons (1992). One-sided nonparametric comparison of treatments with a standard for unequal sample sizes. Journal of Experimental Education **60**, 235-242.

Chakraborti, S. and J.D. Gibbons (1991). One-sided nonparametric comparison of treatments with a standard in the one-way layout. Journal of Quality Technology **23**, 102-106.

Chakraborti, S. and R. Mukerjee (1989). A confidence interval for a measure associated with the comparison of a treatments with a control. South African Statistical Journal. **23**, 219-230.

Chakravarti, I.M., F.C. Leone and J.D. Alanen (1962). Asymptotic relative efficiency of Mood's and Massey's two-sample tests against some parametric alternatives. Annals of Mathematical Statistics **33**, 1375-1383.

Chakravarti, I.M., F.C. Leone and J.D. Alanen (1961). Exact power of some quick tests based on Mood's and Massey's statistics. Bulletin of the Institute of International Statistics **39**, 33e session, Paris, paper no. 110.

Dixon, W.J. (1954). Power under normality of several nonparametric tests. Annals of Mathematical Statistics **25**, 610-614.

Eilbott, J. and J. Nadler (1965). On precedence life testing. Technometrics **7**, 359-377.

Epstein, B. (1955). Comparison of some non-parametric tests against normal alternatives with an application to life testing. Journal of the American Statistical Association **50**, 894-900.

Epstein, B. (1954). Tables for the distribution of the number of exceedances. Annals of Mathematical Statistics **25**, 762-768.

Fligner, M.A. and D.A. Wolfe (1976). Some applications of sample analogs to the probability integral transformation and a coverage property. American Statistician **30**, 78-85.

Gart, J.J. (1963). A median test with sequential applications. Biometrika **50**, 55-62.

Gastwirth, J.L. (1968). The first median test: a two-sided version of the control median test. Journal of the American Statistical Association **63**, 692-706.

Gibbons, J.D. (1976). *Nonparametric methods for quantitative analysis.* Holt, Rinehart and Winston, New York-London.

Gibbons, J.D. and S. Chakraborti (1992). *Nonparametric statistical inference.* 3rd Edition, Marcel Dekker, New York.

Gore A.P., K.S.M. Rao and V. Sahasrabudhe (1986). Median tests for ordered alternatives. Gujarat Statistical Review **XIII**, 55-63.

Gumbel, E.J. and H. von Schelling (1950). The distribution of the number of exceedances. Annals of Mathematical Statistics **21**, 247-262.

Gupta, S.S. (1963). Probability integrals of the multivariate normal and multivariate t. Annals of Mathematical Statistics **34**, 792-826.

Hackl, W. and P. Katzenbeisser (1984). A note on the power of two-sample tests for dispersion based on exceedance statistics. Computational Statistics Quarterly **1**, 333-341.

Haga, T. (1959/1960). A two-sample rank test on location. Annals of the Institute of Statistical Mathematics **11**, 211-219.

Hájek, J. and Z. Sidák (1967). *Theory of Rank Tests.* Academic Press, New York-London.

Harris, L.B. (1952). On a limiting case for the distribution of exceedances, with an application to life-testing. Annals of Mathematical Statistics **23**, 295-298.

Hastings Jr., C. (1955). *Approximations for digital computers.* Princeton University Press, Princeton, New Jersey.

Hemelrijk, J. (1950). Symmetrietoetsen en andere toepassingen van de theorie van Neyman en Pearson. Thesis, University of Amsterdam.

Ismail, H. (1992). Nonparametric Comparison of Treatments with a Control: Procedures Based on the Sign Test. Unpublished Doctoral Dissertation, University of Alabama.

Kao, N. and S. Chakraborti (1994). Small sample null distribution of two nonparametric tests for comparing treatments with a control. Submitted.

Katzenbeisser, W. (1989). The exact power of two-sample location tests based on exceedance statistics against shift alternatives. Mathematische Operationsforschung und Statistik, Series Statistics **20**, 47-54.

Lehmann, E.L. (1975). *Nonparametrics: statistical methods based on ranks.* Holden-Day, San Francisco.

Lehmann, E.L. (1963). Nonparametric confidence intervals for a shift parameter. Annals of Mathematical Statistics **34**, 1507-1512.

Lehmann, E.L. (1959). *Testing Statistical Hypothesis.* John Wiley, New York-London.

Lin, C. H. and S. Sukhatme (1992). On the choice of precedence tests. Communications in Statistics: Theory and Methods **21**, 2949-2968.

Liu, J. (1992). Precedence probabilities and their applications. Communications in Statistics: Theory and Methods **21**, 1667-1682.

Massey, F.J. (1951). A note on a two-sample test. Annals of Mathematical Statistics **22**, 304-306.

Mathisen, H.C. (1943). A method of testing the hypothesis that two samples are from the same population. Annals of Mathematical Statistics **14**, 188-194.

Mood, A.M. (1954). On the asymptotic relative efficiency of certain nonparametric two-sample tests. Annals of Mathematical Statistics **25**, 514-522.

Mood, A.M. (1950). *Introduction to the theory of statistics.* McGraw-Hill Book Company, Inc., New York-London.

Mood, A.M. and F.A. Graybill (1963). *Introduction to the theory of statistics.* McGraw-Hill Book Company, Inc., New York-London.

Moses, E. (1965). Answer to Query 10: Confidence limits from rank tests. Technometrics **7**, 257-260.

Nelson, L.S. (1993). Tests on early failures - the precedence life test. Journal of Quality Technology **25**, 140-149.

Nelson, L.S. (1963). Tables for a precedence life test. Technometrics **5**, 491-499.

Orban, J.J. and D.A. Wolfe (1982). A class of distribution-free tests based on placements. Journal of the American Statistical Association **77**, 666-671.

Pearson, E.S. and H.O. Hartley (1958). *Biometrika Tables for Statisticians,* Volume I. Cambridge University Press.

Pratt, J. W. (1964). Robustness of some procedures for the two-sample location problem. Journal of the American Statistical Association **59**, 665-680.

Rao, K. S. M. and A. P. Gore (1984). Testing against ordered alternatives in the one-way layout. Biometrical Journal **26**, 25-32.

Robertson T., F. T. Wright and R. L. Dykstra (1989). *Order Restricted Statistical Inference.* John-Wiley, New York.

Rosenbaum, S. (1954). Tables for a nonparametric test of location. Annals of Mathematical Statistics **25**, 146-150.

Sarkadi, K. (1957). On the distribution of the number of exceedances. Annals of Mathematical Statistics **28**, 1021-1023.

Schaafsma, W. (1966). *Hypothesis Testing Problems with the Alternative Restricted by a Number of Inequalities*, Noordhoff, Groningen, The Netherlands.

Schlittgen, R. (1979). Use of a median test for a generalized Behrens-Fisher problem. Metrika **26**, 95-103.

Sen, P.K. (1966). On a distribution-free method of estimating asymptotic efficiency of a class of nonparametric tests. Annals of Mathematical Statistics **37**, 1759-1770.

Sen, P.K. (1962). On the role of a class of quantile tests in some multisample nonparametric problems. Calcutta Statistical Association Bulletin **11**, 125-143.

Shorack, R.A. (1967). On the power of precedence life test. Technometrics **9**, 154-158.

Sidák Z. and J. Vondrácek (1957). A simple nonparametric test of difference in location of two populations. (Czech.) Aplikace matematiky **2**, 215-221.

Slivka (1970). A one-sided nonparametric multiple comparison control percentile test: treatments versus control. Biometrika **57**, 431-438.

Sukhatme, S. (1992). Powers of two-sample rank tests under Lehmann alternatives. American Statistician **46**, 212-214.

Thompson, W.R. (1938). Biological applications of normal range and associated significance tests in ignorance of original distribution forms. Annals of Mathematical Statistics **9**, 281-287.

Tsao, C.K. (1954). An extension of Massey's distribution of the maximum deviation of two cumulative step functions. Annals of Mathematical Statistics **25**, 587-592.

Tukey, J.W. (1959). A quick, compact, two-sample test to Duckworth's specifications. Technometrics **1**, 31-48.

Van der Laan, P. (1970 a). Simple distribution-free confidence intervals for location difference, Abstract, Annals of Mathematical Statistics **41**, 331.

Van der Laan, P. (1970 b). Simple distribution-free confidence intervals for a difference in location. Thesis, Eindhoven University of Technology.

Westenberg, J. (1952). A tabulation of the median test with comments and corrections to previous papers. Indagationes Mathematicae **14** (Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen **55**), 10-15.

Westenberg, J. (1950). A tabulation of the median test for unequal samples. Indagationes Mathematicae **12** (Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen **53**), 77-82.

Westenberg, J. (1948). Significance test for median and interquantile range in samples from continuous populations of any form. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen **51**, 252-261.

Wilk, M. B. and R. Gnanadesikan. (1968). Probability plotting methods for the analysis of data. Biometrika **55**, 1-17.

Wilks, S. S. (1963). *Mathematical Statistics.* John Wiley, New York.

Young, D. H. (1973). A note on some asymptotic properties of the precedence test and applications to a selection problem concerning quantiles. Sankhya, Series B **35**, 35-44.

Table 3.1

Observed probability of accepting $H_0(d=0)$ based on 200 pairs of samples, each of size ten.

	rank		exceedance			maximum deviation		
$d = \lvert\frac{\mu_1-\mu_2}{\sigma}\rvert$	sum	run	$r=1$	$r=2$	$r=3$	$r=3$	$r=6$	$r=10$
0	.935	.965	.95	.96	.96	.955	.945	.945
1	.485	.795	.655	.65	.60	.575	.555	.555
2	.015	.275	.16	.12	.10	.065	.045	.045
3	0	.02	.025	0	0	0	0	0

Table 5.1. Distribution-free lower confidence bounds D_{ji}, denoted by $j-i$, for shift (selection based on Normal shift alternatives) with confidence level $1-\alpha = .01$ and the values of: 1-confidence coefficient.

$\Pr[\underline{y}_{(j)} - \underline{x}_{(i)} < \nu] \geqslant \cdot 95$

m \ n	3	4	5	6	7	8	9	10	11	12	13	14	15
3	1—3 ·05000	1—3 ·02857	1—3 ·01786	2—3 ·04762	2—3 ·03333	2—3 ·02424	3—3 ·04545	3—3 ·03497	3—3 ·02747	4—3 ·04396	4—3 ·03571	4—3 ·02941	5—3 ·04289
4	1—4 ·02857	1—4 ·01429	1—3 ·04762	1—3 ·03333	3—4 ·04545	3—4 ·03030	4—4 ·04895	2—3 ·04096	2—3 ·03297	1—2 ·05000	1—2 ·04412	3—3 ·04412	3—3 ·03741
5	1—5 ·01786	2—5 ·04762	1—4 ·02381	3—5 ·04545	2—4 ·04545	4—5 ·04351	2—4 ·02298	3—4 ·04695	3—4 ·03571	6—5 ·04072	4—4 ·04739	4—4 ·03793	4—4 ·03070
6	1—5 ·04762	2—6 ·03333	1—4 ·04545	2—5 ·04004	4—6 ·04895	4—6 ·02797	2—4 ·04695	2—4 ·03571	4—5 ·04299	7—6 ·04977	5—5 ·04954	5—5 ·03793	3—4 ·03070
7	1—6 ·03333	1—5 ·04545	2—6 ·04545	1—4 ·04895	4—7 ·03497	2—5 ·03497	2—5 ·02448	4—6 ·03640	2—4 ·04739	5—6 ·03989	5—6 ·02864	4—5 ·04076	4—5 ·03223
8	1—7 ·02424	1—6 ·03030	1—5 ·04351	1—5 ·02797	3—7 ·03497	1—4 ·03846	3—6 ·04447	3—6 ·03065	5—7 ·03711	4—6 ·03989	3—5 ·04076	5—6 ·04799	5—6 ·03668
9	1—7 ·04545	1—6 ·04895	2—8 ·02298	3—8 ·04695	3—8 ·02448	3—7 ·04447	3—7 ·02834	2—5 ·04954	3—6 ·03989	6—8 ·03746	5—7 ·04025	4—6 ·04158	6—7 ·04469
10	1—8 ·03497	2—9 ·04096	2—8 ·04695	3—9 ·03571	2—7 ·03640	3—8 ·03065	5—9 ·04954	3—7 ·03489	6—9 ·04257	4—7 ·04561	3—6 ·03668	6—8 ·04025	5—7 ·04158
11	1—9 ·02747	2—10 ·03297	2—9 ·03571	2—8 ·04299	4—10 ·04739	2—7 ·03711	4—9 ·03989	2—6 ·04257	4—8 ·04305	7—10 ·04469	5—8 ·04977	4—7 ·04158	9—10 ·04943
12	1—9 ·04396	3—12 ·05000	1—7 ·04072	1—6 ·04977	2—8 ·03989	3—9 ·03989	2—7 ·03746	4—9 ·04561	2—6 ·04469	5—9 ·049[illegible]	4—8 ·03581	7—10 ·04202	4—7 ·04928
13	1—10 ·03571	3—13 ·04412	2—10 ·04739	2—9 ·04954	2—9 ·02864	4—11 ·04076	3—9 ·04025	5—11 ·03668	4—9 ·04977	5—10 ·03581	7—11 ·04842	5—9 ·04123	8—11 ·04343
14	1—11 ·02941	2—12 ·04412	2—11 ·03793	2—10 ·03793	3—11 ·04076	3—10 ·04799	4—11 ·04158	3—9 ·04025	5—11 ·04158	3—8 ·04202	5—10 ·04123	5—10 ·02849	5—9 ·04756
15	1—11 ·04289	2—13 ·03741	2—12 ·03070	3—13 ·03070	3—12 ·03223	3—11 ·03668	3—10 ·04469	4—11 ·04158	2—7 ·04943	6—12 ·04928	3—8 ·04343	6—11 ·04756	5—10 ·03280

Table 5.2. Distribution-free lower confidence bounds D_{ji}, denoted by $j-i$, for shift (selection based on Normal shift alternatives) with confidence level $1-\alpha$ = .05 and the values of: 1-confidence coefficient.

$\Pr[\underline{y}_{(j)} - \underline{x}_{(i)} < \nu] \geqslant \cdot 99$

m \ n	3	4	5	6	7	8	9	10	11	12	13	14	15
3					1–3 ·00833	1–3 ·00606	1–3 ·00455	1–3 ·00350	1–3 ·00275	2–3 ·00879	2–3 ·00714	2–3 ·00588	2–3 ·00490
4			1–4 ·00794	1–4 ·00476	1–4 ·00303	1–4 ·00202	2–4 ·00699	2–4 ·00500	1–3 ·00879	3–4 ·00824	3–4 ·00630	3–4 ·00490	4–4 ·00903
5		1–5 ·00794	1–5 ·00397	1–5 ·00216	2–5 ·00758	1–4 ·00699	1–4 ·00500	3–5 ·00699	3–5 ·00481	2–4 ·00986	2–4 ·00770	2–4 ·00611	5–5 ·00813
6		1–6 ·00476	1–6 ·00216	1–5 ·00758	1–5 ·00466	3–6 ·00932	3–6 ·00559	2–5 ·00762	2–5 ·00541	2–5 ·00393	3–5 ·00955	3–5 ·00722	6–6 ·00851
7	1–7 ·00833	1–7 ·00303	1–6 ·00758	2–7 ·00466	1–6 ·00233	2–6 ·00886	2–6 ·00559	4–7 ·00617	3–6 ·00905	2–5 ·00955	2–5 ·00722	4–6 ·00898	4–6 ·00661
8	1–8 ·00606	1–8 ·00202	2–8 ·00699	1–6 ·00932	2–7 ·00886	2–7 ·00505	1–5 ·00905	2–6 ·00905	2–6 ·00627	4–7 ·00988	4–7 ·00671	3–6 ·00832	5–7 ·00841
9	1–9 ·00455	1–8 ·00699	2–9 ·00500	1–7 ·00559	2–8 ·00559	4–9 ·00905	2–7 ·00761	2–7 ·00488	4–8 ·00917	3–7 ·00845	3–7 ·00588	5–8 ·00693	4–7 ·00884
10	1–10 ·00350	1–9 ·00500	1–8 ·00699	2–9 ·00762	1–7 ·00617	3–9 ·00905	3–9 ·00488	2–7 ·00988	3–8 ·00733	3–8 ·00478	5–9 ·00650	3–7 ·00884	3–7 ·00650
11	1–11 ·00275	2–11 ·00879	1–9 ·00481	2–10 ·00541	2–9 ·00905	3–10 ·00627	2–8 ·00917	3–9 ·00733	3–9 ·00446	4–9 ·00950	3–8 ·00693	6–10 ·00704	5–9 ·00771
12	1–11 ·00879	1–10 ·00824	2–11 ·00986	2–11 ·00393	3–11 ·00955	2–9 ·00988	3–10 ·00845	3–10 ·00478	3–9 ·00950	4–10 ·00614	2–7 ·00998	4–9 ·00893	4–9 ·00617
13	1–12 ·00714	1–11 ·00630	2–12 ·00770	2–11 ·00955	3–12 ·00722	2–10 ·00671	3–11 ·00588	2–9 ·00650	4–11 ·00693	6–12 ·00998	4–10 ·00847	7–12 ·00876	6–11 ·00830
14	1–13 ·00588	1–12 ·00490	2–13 ·00611	2–12 ·00722	2–11 ·00898	3–12 ·00832	2–10 ·00693	4–12 ·00884	2–9 ·00704	4–11 ·00893	2–8 ·00876	3–9 ·00915	[illegible]–10 ·00729
15	1–14 ·00490	1–12 ·00903	1–11 ·00813	1–10 ·00851	2–12 ·00661	2–11 ·00841	3–12 ·00884	4–13 ·00650	3–11 ·00771	4–12 ·00617	3–10 ·00830	5–12 ·00729	5–12 ·00461

Table 5.3.

ARE of the D_{ji} test relative to Wilcoxon-rank-sum test for various distributions; ρ being the shape parameter of the Gamma distribution.

Uniform	.333	Gamma	
Normal	.667	$\rho = 4$	.601
Logistic	.750	$\rho = 10$	.642
Cauchy	1.333	$\rho = 40$	.661
Double Exponential	1.333		

Table 5.4.

ARE of the D_{ji} test relative to the two sample t-test for various distributions; ρ being the shape parameter of the Gamma distribution.

Uniform	.333	Gamma	
Normal	$\frac{2}{\pi} = .637$	$\rho = 4$	.704
Logistic	$\frac{\pi^2}{12} = .822$	$\rho = 10$	.662
Double Exponential	2	$\rho = 40$	.643

Table 6.1

Values of the ratio of the expected lengths of the selected one-sided confidence intervals and the one-sided confidence intervals corresponding with the two-sample t-test in the case of normal distributions with the same confidence coefficients (confidence level $1 - \alpha = .95$)

m \ n	3	4	5	6	7	8	9	10	11	12	13	14	15
3	1·03												
4	1·05	1·06											
5	1·06	1·17	1·16										
6	1·13	1·18	1·22	1·17									
7	1·13	1·20	1·17	1·29	1·30								
8	1·13	1·19	1·26	1·27	1·20	1·35							
9	1·16	1·23	1·17	1·21	1·20	1·20	1·20						
10	1·16	1·18	1·18	1·21	1·21	1·20	1·27	1·21					
11	1·16	1·18	1·18	1·20	1·26	1·24	1·22	1·28	1·21				
12	1·18	1·33	1·29	1·35	1·22	1·21	1·26	1·22	1·31	1·22			
13	1·18	1·34	1·19	1·21	1·22	1·23	1·22	1·24	1·22	1·22	1·27		
14	1·18	1·18	1·19	1·21	1·21	1·21	1·22	1·23	1·23	1·26	1·22	1·22	
15	1·19	1·18	1·19	1·21	1·21	1·21	1·22	1·22	1·33	1·24	1·27	1·23	1·23