DEPARTMENT OF STATISTICS
The University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT NO. 167
July 1968

PROPERTIES OF A RANK TEST OF
CRONHOLM AND REVUSKY

By
G. K. Bhattacharyya
and
Richard A. Johnson
University of Wisconsin, Madison

This research was supported by the Office of Naval Research under contract Nonr 1202(17), Project No. 042-222 and by the Wisconsin Alumni Research Foundation through grants 67-08500-1 and 67-45435-1 for computer usage.
PROPERTIES OF A RANK TEST OF CRONHOLM AND REVUSKY

By

G. K. Bhattacharyya and Richard A. Johnson
University of Wisconsin, Madison

SUMMARY

Cronholm and Revusky (1965) have proposed a rank test based on independent subexperiments for the comparison of a treatment effect with a control when the treatment is destructive in nature but the control produces at most a transient effect. The present work enhances the utility of the test in practical situations by providing a thorough study of its performance under two sampling situations. With the aid of a computer, the exact small sample power of the test is evaluated for some important alternatives. A modification of the test is proposed for the situation where numerical measurements are available rather than merely the ranks within each stage. The modification is found to substantially improve the power. Both the tests are shown to be unbiased. Finally, the Pitman asymptotic efficiencies are obtained and comparisons are made with the appropriate Wilcoxon tests.
1. INTRODUCTION

We are concerned with the problem of testing the equality of a treatment and a control effect in the situation where an application of the treatment to an experimental unit makes the unit unsuitable for further experimentation whereas the control can be replicated on a unit after suitable waiting periods. With \( n \) homogeneous units available, the Cronholm-Revusky sampling scheme effectively increases the number of observations beyond \( n \) through independent subexperiments. Their procedure is to first apply the treatment to a unit selected at random and the control to the remaining \((n-1)\) units. The unit receiving the treatment is then discarded and a suitable waiting period is allowed for the effect of the control to disappear. In the second stage, a unit drawn at random from the \((n-1)\) receives the treatment and the remaining \((n-2)\) receive the control. Thus in the \( i^{th} \) stage, one unit selected at random from the available \((n-i+1)\) receives the treatment while the other \((n-i)\) receive the control. The process continues until all units have received the treatment.

Depending on the observable responses, two types of data could arise from this \( n \)-stage procedure.

(a) If the responses are directly measurable, the data would consist of numerical measurements \( \{(X_{i1}, X_{i2}, \ldots, X_{i, n-1}, Y_i); i=1, 2, \ldots, n\} \) where \( Y_i \) is the treatment response at the \( i^{th} \) stage and \( X_{i1}, \ldots, X_{i, n-1} \) are the control responses at the same stage.

(b) In certain circumstances where the responses are not directly measurable, it may be possible to rank the responses
within a given stage. The data would then be the set of ranks \( \{R_i; i=1,2,\ldots,n\} \) where \( R_i \) is the rank of \( Y_i \) among \( \{Y_i,X_{i1},\ldots,X_{in},n-i\} \). Here the \( X \)'s and the \( Y \)'s are unobservable.

The latter situation occurs, for example, in psychometric studies which employ preference scales. The difference between the two cases is important since comparisons may be made between stages when the numerical scores are available. This point seems to have been previously overlooked.

To formally specify the testing problem, we let \( F(x) \) and \( G(x) \) denote the cumulative distribution functions (cdf's) of the control response \( X \) and the treatment response \( Y \) respectively where the unknown \( F \) and \( G \) are assumed to be continuous. The problem is to test \( H_0: F=G \) against the alternative that \( Y \) is stochastically larger than \( X \), that is, \( G(x) \geq F(x) \) for all \( x \) with strict inequality for some \( x \).

The Cronholm-Revusky test of level \( \alpha \) is given by the critical function

\[
\phi(T_n) = \begin{cases} 
1, & \text{if } T_n > C_n \\
\gamma_n, & \text{if } T_n = C_n \\
0, & \text{otherwise}
\end{cases} \tag{1.1}
\]

where

\[
T_n = \sum_{i=1}^{n} R_i - n \tag{1.2}
\]

and \( C_n \) and \( \gamma_n \) are such that \( E_{H_0}(\phi(T_n)) = \alpha \). This test is distribution-free and is applicable in both situations (a) and (b).

A table of exact tail probabilities is given by Cronholm and Revusky (1965) for \( n \leq 12 \) and a normal approximation is also presented. Further, a rather vague criterion called the "Sensitivity Index" is used by the authors
to compare their test with the single stage Wilcoxon test based on equal sample sizes of \( n/2 \) for both treatment and control. Bennett (1967) studies the limiting form of the cumulant generating function of \( T_n \) under the null hypothesis and under a logit-type alternative.

The present study is motivated by the need for more information about the behavior of the test \( T_n \) under alternatives than is available in the works mentioned above. The exact power of the test for shift alternatives in normal, uniform and exponential distributions as well as for the Lehmann alternatives is evaluated with the help of a recursive scheme. Also the Pitman asymptotic efficiency with respect to the appropriate Wilcoxon test is derived. The Wilcoxon test, considered by Cronholm and Revusky (1965) for comparison, makes sense only in the case (b). In the case (a), the comparable Wilcoxon test should be the one based on as many treatment and control replications as are involved in \( T_n \). This point is elaborated in Section 5.
2. SMALL SAMPLE PROPERTIES AND EXACT POWER

For notational simplicity, we make the identification \( U_i = R_{n-i+1}^{-1} \). Note that \( U_i \) is the number of control responses at the \((n-i+1)\)st stage which are less than \( Y_{n-i+1} \). Thus the suffix of \( U \) indicates the total number of observations with which the treatment response is compared at each stage. In this notation, \( T_n = \sum_{i=2}^{n} U_i \) and independent randomization at each stage helps ensure that the summands are independent. This fact makes the development of the distribution theory and evaluation of power substantially simpler for the present test than for the usual single stage two sample rank tests.

Theoretically ties are impossible because of the assumption of continuous cdf's. If they occur in practice, they could be handled by the standard method of using midranks.

Setting \( t_{ij} = 1 \) if \( X_{ij} < Y_i \) and \( t_{ij} = 0 \) otherwise, we have

\[
U_i = \sum_{j=1}^{i-1} t_{n-i+1,j} \quad \text{for } i=2, \ldots, n.
\]

The conditional distribution of \( U_i \) given \( Y_{n-i+1} = y \) is the binomial distribution with \((i-1)\) trials and success probability \( F(y) \). Thus the unconditional distribution is given by

\[
P(U_i = k) = P_{ki} = \binom{i-1}{k} \int F^{k}(y)[1-F(y)]^{i-k-1}dG(y)
\]

for \( k = 0, 1, \ldots, i-1 \). (2.1)

The mean and variance of \( T_n \), which are easily computed in terms of the indicator variables \( t_{ij} \), are given by

\[
\mu_n = E(T_n) = \binom{n}{2} \rho(F,G)
\]

(2.2)

\[
\sigma_n^2 = Var(T_n) = \binom{n}{2} [\rho(F,G) - \rho^2(F,G)] + 2\binom{n}{3} \xi^2(F,G)
\]
where

\[ \rho(F, G) = \int F(x) dG(x), \]

\[ \zeta^2(F, G) = \int F^2(x) dG(x) - \rho^2(F, G) = \text{Var}_G[F(X)]. \] (2.3)

Since the \( U_i \) are independent Mann-Whitney U statistics, the third and fourth moments of \( T_n \) can be computed from the results in Sundrum (1954).

From (2.1) the probability generating function (pgf) of \( U_i \) is given by

\[ g_1(t) = \int [1 + (t-1)F(y)]^{i-1} dG(y) \] (2.4)

and hence that of \( T_n \) is the product \( g(t) = \prod_{i=2}^{n} g_i(t) \). Under the null hypothesis \( F = G \), the pgf of \( T_n \) simplifies to

\[ g_0(t) = [n!(t-1)^{n-1}]^{-1} \prod_{i=2}^{n} (t^i-1) \] (2.5)

and this expression was used by Cronholm and Revusky (1965) to tabulate the null distribution of \( T_n \) for \( n \leq 12 \).

We mention here an interesting relation between the null distribution of \( T_n \) and the distribution of the well-known Kendall's \( \tau \)-statistic under independence. Let \( (V_i, W_i) \), \( i = 1, 2, \ldots, n \) be a random sample from a continuous bivariate distribution. Arranging these vectors in increasing order of the first coordinate, we have \( (V(i), W[i]) \), \( i = 1, 2, \ldots, n \) where \( V(1) < V(2) < \ldots < V(n) \) and \( W[i] \) is the value of \( W \) associated with the \( i \)th smallest value of \( V \). Then disregarding constant multipliers, Kendall's \( \tau \)-statistic can be expressed as

\[ \tau = \sum_{1 \leq i \leq j \leq n} e_{ij} = \sum_{i=1}^{n} I(i) \] (2.6)
where \( e_{ij} = 1 \) \((0)\) according as \( W_{[i]} \leq (>) W_{[j]} \) and \( \ell_{(1)} \) is the rank of \( W_{[1]} \) among \( \{ W_{[1]}, \ldots, W_{[i]} \} \). When \( V \) and \( W \) are independent, the random variables \( \ell_{(1)}, \ldots, \ell_{(n)} \) are independently distributed and \( \ell_{(i)} \) has uniform distribution over the set of integers \( \{1, 2, \ldots, i\} \) [see Brandorff-Nielson (1963)]. Hence the pgf of \( \tau - n \) is precisely the same as \( g_{O}(t) \) given in (2.5). An elaborate table of significance points of \( \tau \) for sample sizes up to \( n = 40 \) has been prepared by Kaarsemaker and Wijngaarden (1953). Due to the above correspondence, the same table can be used to carry out the Cronholm-Revusky test.

For any specific alternative \((F, G)\), set \( P_{n}(\ell) = P_{F, G}(T_{n} = \ell) \). Since

\[
T_{n} = T_{n-1} + U_{n},
\]

where the two terms on the right hand side are independent, we have the simple recursive relation

\[
P_{n}(\ell) = \sum_{k=0}^{n-1} P_{n-1}(\ell-k) p_{kn}
\]

(2.7)

with boundary conditions \( P_{2}(\ell) = p_{2} \) for \( \ell = 0, 1 \) and zero otherwise. Consequently, once the basic probabilities \( p_{k1} \) in (2.1) are available, the distribution of \( T_{n} \) and also the power of the test can be determined recursively by (2.7). The following two important classes of nonparametric alternatives are considered.

1. Shift alternative: \( G(x) = F(x - \Delta), \Delta > 0 \)

2. Lehmann alternative: \( G(x) = F^{\theta}(x), \theta > 1 \).

Both of these are special cases of \( Y \) being stochastically larger than \( X \). When \( \theta \) is an integer, (2) has the interpretation that the distribution of \( Y \) is the same as that of \( \max \{X_1, X_2, \ldots, X_0\} \) where \( X_1, \ldots, X_0 \) are i.i.d \( F(x) \). The shift alternative (1), on the other hand, is a model for additive
treatment effect.

For shift alternatives, let \( p_{k1}^{(1)}(\Delta), \ p_{k1}^{(2)}(\Delta) \) and \( p_{k1}^{(3)}(\Delta) \) denote the probability (2.1) specialized to the normal: \( F(x) = \Phi(x) \), the uniform: \( F(x) = x, \ 0 < x < 1 \) and the exponential: \( F(x) = 1 - e^{-x}, \ x > 0 \), respectively. The values of \( p_{k1}^{(1)}(\Delta) \), for some selected \( \Delta \), are read directly from Milton (1965) upon identifying the appropriate rank order for the case \( m = 1 - l \) and \( N = i \). Using these along with the recursive scheme (2.7), the power is computed on the B5500 for sample sizes \( n = 2(l)12 \) and alternatives \( \Delta = .4, .8, 1.5, 2.0 \) and 3.0. These are exhibited in Table 1. Nonrandomized significance levels are chosen close to 1% and 5% and these are entered in the column \( \Delta = 0 \).

By straightforward integration, one obtains

\[
p_{k1}^{(2)}(\Delta) = i^{-1}[1 - I(k+1, i-k, \Delta)], \quad \text{if } 0 \leq k \leq i-1
\]
\[
= i^{-1}[1 + i\Delta - \Delta^i], \quad \text{if } k = i-1
\]

\[
p_{k1}^{(3)}(\Delta) = i^{-1}e^{\Delta}[1 - I(k+1, i-k, 1-e^{-\Delta})], \quad \text{if } 0 \leq k \leq i-1;
\]

(2.8)

(2.9)

where \( I \) is the incomplete beta integral given by

\[
I(a, b, y) = \int_0^y x^{a-1}(1-x)^{b-1}dx/\int_0^1 x^{a-1}(1-x)^{b-1}dx.
\]

Note that for \( k < i-1 \), (2.9) is related to (2.8) by

\[
(1-\Delta)p_{k1}^{(3)}[-\log(1-\Delta)] = p_{k1}^{(2)}(\Delta).
\]

(2.10)

Employing Pearson's tables (1956) to first evaluate \( p^{(2)} \) and \( p^{(3)} \), we compute the power by again using the recursive scheme (2.7). Table 2 presents the power for sample sizes \( n = 2(l)10 \) and for some selected values of \( \Delta \).
Finally, for the Lehman alternative \( G(x) = F^{\theta}(x) \), the probability (2.1) simplifies to

\[
p_{ki}^{(0)}(\theta) = \frac{\theta^{k-1}}{\theta^{i-1}} \quad , \quad 0 \leq k \leq i-1 \tag{2.11}
\]

when \( \theta \) is an integer. For \( \theta = 2 \) and 3, power is computed by using (2.11) and the same procedure as in the previous cases. The results are given in Table 2. The entries in Table 1 and 2 are accurate up to the last decimal place retained. Some checks have been made by direct computation.

We complete this section by showing that the Cronholm-Revensky test is unbiased for a wide class of alternatives.

**Theorem 2.1.** The one sided level \( \alpha \) test \( \phi(T_n) \) given in (1.1) is unbiased for all distributions \( F(x) \) and \( G(y) \) such that \( Y \) is stochastically larger than \( X \).

**Proof.** By definition, \( E_{H_0}[\phi(T_n)] = \alpha \). Also \( T_n = T_n(x_{ij}, y_{i'}; 1 \leq j \leq n-i, i=1,2,\ldots,n) \) is a function of the \( X \) and \( Y \) values coming from all the stages. One can easily verify that \( T_n \) is nondecreasing in each \( Y_i \) when the other arguments are held fixed. Hence the conditions of Lemma 2 on page 187 of Lehmann (1959) are satisfied and the conclusion follows.
Table 1. Power of $T_n$ for normal shift alternatives.

<table>
<thead>
<tr>
<th>n</th>
<th>0</th>
<th>.4</th>
<th>.8</th>
<th>1.5</th>
<th>2.0</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.5000</td>
<td>.6114</td>
<td>.7142</td>
<td>.8556</td>
<td>.9214</td>
<td>.9831</td>
</tr>
<tr>
<td>3</td>
<td>.1667</td>
<td>.2763</td>
<td>.4103</td>
<td>.6552</td>
<td>.7977</td>
<td>.9524</td>
</tr>
<tr>
<td>4</td>
<td>.0417</td>
<td>.1002</td>
<td>.2003</td>
<td>.4599</td>
<td>.6563</td>
<td>.9108</td>
</tr>
<tr>
<td>5</td>
<td>.0083</td>
<td>.0306</td>
<td>.0858</td>
<td>.3002</td>
<td>.5171</td>
<td>.8610</td>
</tr>
<tr>
<td></td>
<td>.0417</td>
<td>.1172</td>
<td>.2553</td>
<td>.5935</td>
<td>.8004</td>
<td>.9751</td>
</tr>
<tr>
<td>6</td>
<td>.0083</td>
<td>.0372</td>
<td>.1171</td>
<td>.4232</td>
<td>.6830</td>
<td>.9551</td>
</tr>
<tr>
<td></td>
<td>.0681</td>
<td>.1948</td>
<td>.4044</td>
<td>.7888</td>
<td>.9343</td>
<td>.9973</td>
</tr>
<tr>
<td>7</td>
<td>.0151</td>
<td>.0703</td>
<td>.2135</td>
<td>.6358</td>
<td>.8673</td>
<td>.9935</td>
</tr>
<tr>
<td></td>
<td>.0681</td>
<td>.2162</td>
<td>.4636</td>
<td>.8603</td>
<td>.9697</td>
<td>.9995</td>
</tr>
<tr>
<td>8</td>
<td>.0156</td>
<td>.0824</td>
<td>.2616</td>
<td>.7338</td>
<td>.9295</td>
<td>.9985</td>
</tr>
<tr>
<td></td>
<td>.0543</td>
<td>.2032</td>
<td>.4726</td>
<td>.8887</td>
<td>.9819</td>
<td>.9999</td>
</tr>
<tr>
<td>9</td>
<td>.0124</td>
<td>.0787</td>
<td>.2746</td>
<td>.7802</td>
<td>.9549</td>
<td>.9996</td>
</tr>
<tr>
<td></td>
<td>.0597</td>
<td>.2367</td>
<td>.5439</td>
<td>.9349</td>
<td>.9932</td>
<td>1.0</td>
</tr>
<tr>
<td>10</td>
<td>.0083</td>
<td>.0665</td>
<td>.2651</td>
<td>.8008</td>
<td>.9667</td>
<td>.9998</td>
</tr>
<tr>
<td></td>
<td>.0542</td>
<td>.2405</td>
<td>.5726</td>
<td>.9544</td>
<td>.9967</td>
<td>1.0</td>
</tr>
<tr>
<td>11</td>
<td>.0083</td>
<td>.0741</td>
<td>.3042</td>
<td>.8555</td>
<td>.9830</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>.0433</td>
<td>.2249</td>
<td>.5747</td>
<td>.9632</td>
<td>.9980</td>
<td>1.0</td>
</tr>
<tr>
<td>12</td>
<td>.0105</td>
<td>.0967</td>
<td>.3783</td>
<td>.9134</td>
<td>.9936</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>.0432</td>
<td>.2415</td>
<td>.6169</td>
<td>.9768</td>
<td>.9992</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Table 2. Power of $T_n$ for Lehmann alternatives and shift alternatives in uniform and exponential distributions.

<table>
<thead>
<tr>
<th>n</th>
<th>Lehmann alternative</th>
<th>Uniform shift</th>
<th>Exponential shift</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0=2</td>
<td>0=3</td>
</tr>
<tr>
<td>2</td>
<td>.5000</td>
<td>.6667</td>
<td>.7500</td>
</tr>
<tr>
<td>3</td>
<td>.1667</td>
<td>.3333</td>
<td>.4500</td>
</tr>
<tr>
<td>4</td>
<td>.0417</td>
<td>.1333</td>
<td>.2250</td>
</tr>
<tr>
<td>5</td>
<td>.0083</td>
<td>.0444</td>
<td>.0964</td>
</tr>
<tr>
<td>5</td>
<td>.0417</td>
<td>.1652</td>
<td>.2989</td>
</tr>
<tr>
<td>6</td>
<td>.0083</td>
<td>.0578</td>
<td>.1379</td>
</tr>
<tr>
<td>6</td>
<td>.0681</td>
<td>.2810</td>
<td>.4873</td>
</tr>
<tr>
<td>7</td>
<td>.0151</td>
<td>.1141</td>
<td>.2662</td>
</tr>
<tr>
<td>7</td>
<td>.0681</td>
<td>.3230</td>
<td>.5675</td>
</tr>
<tr>
<td>8</td>
<td>.0156</td>
<td>.1402</td>
<td>.3366</td>
</tr>
<tr>
<td>8</td>
<td>.0543</td>
<td>.3185</td>
<td>.5895</td>
</tr>
<tr>
<td>9</td>
<td>.0124</td>
<td>.1414</td>
<td>.3626</td>
</tr>
<tr>
<td>9</td>
<td>.0597</td>
<td>.3768</td>
<td>.6774</td>
</tr>
<tr>
<td>10</td>
<td>.0083</td>
<td>.1273</td>
<td>.3594</td>
</tr>
<tr>
<td>10</td>
<td>.0542</td>
<td>.3947</td>
<td>.7167</td>
</tr>
<tr>
<td>11</td>
<td>.0083</td>
<td>.1474</td>
<td>.4194</td>
</tr>
<tr>
<td>11</td>
<td>.0433</td>
<td>.3856</td>
<td>.7282</td>
</tr>
<tr>
<td>12</td>
<td>.0105</td>
<td>.1951</td>
<td>.5219</td>
</tr>
<tr>
<td>12</td>
<td>.0432</td>
<td>.4203</td>
<td>.7766</td>
</tr>
</tbody>
</table>
3. A MODIFICATION OF $T_n$

We consider here a modification of the Cronholm-Revusky test $\phi(T_n)$ for the situation (a). In the original test, the number of control responses varies from 0 to $n-1$ over the different stages. When numerical measurements are available, one can distribute the control replications more evenly with respect to the treatment observations. Consider first the case where $n$ is odd, say $n=2s+1$. Divide at random the set of $\binom{n}{2}$ control responses $\{X_{ij}\}$ into $n$ sets of $s$ observations each $\{X_{i1}^*, \ldots, X_{is}^*\}$, $i=1, 2, \ldots, n$. Let $U_i^*$ equal the number of observations in the $i$th set which are less than $Y_i$. The modified test statistic is defined as

$$T_n^* = \sum_{i=1}^{n} U_i^*$$  \hspace{1cm} (3.1)

and again the null hypothesis is to be rejected for large values of $T_n^*$. Compared to $T_n$, the probability distribution of $T_n^*$ has the added simplicity that the summands $U_i^*$ in (3.1) are not only independent but also identically distributed. The pgf of $T_n^*$ under an arbitrary alternative $(F, G)$ is given by

$$g^*(t) = \{ \int [1 + (t-1)F(y)]^S dG(y) \}^n$$  \hspace{1cm} (3.2)

and under $H_0$ this reduces to

$$g_{0}^*(t) = [(t^{s+1}-1)/s(t-1)]^n.$$  \hspace{1cm} (3.3)

When $n$ is even ($=2s$), the form of the modified test is self evident. The first $s$ sets constructed will each contain $(s-1)$ observations and the others will contain $s$. In this case, $T_n^* = \sum_{i=1}^{s} U_i^* + \sum_{i=s+1}^{n} U_i^*$ and each of the two terms is a sum of i.i.d components.
Employing the rank order probabilities in Milton (1965), the power of the test $T^*_n$ for normal shift alternatives is evaluated and the results are exhibited in Table 3. For ease of computation, only odd sample sizes are considered. The recursive scheme for this modified test is easier to handle than that for $T_n^*$. In the next two sections, it is shown that this modification improves upon the asymptotic efficiency as well as the small sample exact power. Note that $T^*_n$ is also unbiased for all stochastically larger alternatives and the proof is essentially the same as that of Theorem 2.1.

<table>
<thead>
<tr>
<th>n</th>
<th>0</th>
<th>.4</th>
<th>.8</th>
<th>1.5</th>
<th>2.0</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.1250</td>
<td>.2285</td>
<td>.3643</td>
<td>.6263</td>
<td>.7821</td>
<td>.9500</td>
</tr>
<tr>
<td>5</td>
<td>.0247</td>
<td>.0853</td>
<td>.2147</td>
<td>.5721</td>
<td>.7987</td>
<td>.9790</td>
</tr>
<tr>
<td>5</td>
<td>.0864</td>
<td>.2270</td>
<td>.4423</td>
<td>.8109</td>
<td>.9437</td>
<td>.9982</td>
</tr>
<tr>
<td>7</td>
<td>.0073</td>
<td>.0469</td>
<td>.1769</td>
<td>.6313</td>
<td>.8834</td>
<td>.9972</td>
</tr>
<tr>
<td>7</td>
<td>.0449</td>
<td>.1818</td>
<td>.4473</td>
<td>.8828</td>
<td>.9820</td>
<td>.9999</td>
</tr>
<tr>
<td>9</td>
<td>.0114</td>
<td>.0894</td>
<td>.3344</td>
<td>.8714</td>
<td>.9863</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>.0381</td>
<td>.2033</td>
<td>.5406</td>
<td>.9551</td>
<td>.9974</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>.0100</td>
<td>.1047</td>
<td>.4215</td>
<td>.9461</td>
<td>.9978</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>.0562</td>
<td>.3082</td>
<td>.7173</td>
<td>.9923</td>
<td>.9999</td>
<td>1.00</td>
</tr>
</tbody>
</table>
4. THE ASYMPTOTIC THEORY

In order to obtain a large sample approximation to power, we first derive the limiting distribution of \( T_n \) when the random variables \( \{X_{ij}; \quad 1 \leq j \leq n-i, \quad 1 \leq i \leq n-1\} \) are i.i.d. \( F_n(x) \) and the random variables \( \{X_i; \quad 1 \leq i \leq n\} \) are i.i.d. \( G_n(y) \). For generality, the continuous cdf's \( F_n \) and \( G_n \) are allowed to vary with \( n \). The mean \( \mu_n \) and variance \( s_n^2 \) of \( T_n \) are given by (2.2) with \( F \) and \( G \) replaced by \( F_n \) and \( G_n \) respectively. Further, let
\[
\mathcal{L}(\cdot | F_n, G_n) \text{ denote the law of } (\cdot) \text{ under } (F_n, G_n) \text{ and let } \mathcal{N}(a, b) \text{ denote the normal law with mean } a \text{ and variance } b.
\]

Bennett (1967) showed limiting normality for \( T_n \) under both \( H_0 \) and a logit type alternative by the method of convergence of moments. The following theorem extends his results to general alternatives \( (F_n, G_n) \) and also establishes the uniformity of the convergence over a certain class of alternatives.

**Theorem 3.1.** Let \( \zeta \) be given by (2.3). If the sequence \( \{(F_n, G_n)\} \) satisfies \( n \xi^2(F_n, G_n) \to \infty \) as \( n \to \infty \), then
\[
\lim_{n \to \infty} \mathcal{L}\left( \frac{T_n - \mu_n}{s_n} \big| F_n, G_n \right) = \mathcal{N}(0, 1) .
\]

Further, if there exists an \( \epsilon > 0 \) such that \( \zeta(F_n, G_n) \geq \epsilon \) for all \( n \), the above convergence is uniform in \( \{(F_n, G_n)\} \).

**Proof.** Let \( H_{ni}(x) \) denote the cdf of the random variable \( Z_{ni} = [U_i - (i-1) \rho(F_n, G_n)]/s_n, \quad 2 \leq i \leq n, \quad n=2,3,\ldots \) where \( \rho \) is given by (2.3). For an arbitrary \( \epsilon > 0 \), define
\[
\eta_n(\epsilon) = \sum_{i=2}^{n} \int_{|x| \leq \epsilon} x^2 dH_{ni}(x) .
\]

(4.2)
Since $0 \leq U_i \leq (i-1)$, we have $|Z_{ni}| < n/s_n$ and this bound tends to zero as $n \to \infty$ under the assumption that $n \zeta^2(F_n, G_n) \to \infty$. Thus there exists an integer $n_o(\epsilon)$ such that $q_n(\epsilon) = 0$ for all $n \geq n_o(\epsilon)$ and (4.1) then follows from the Lindeberg-Feller central limit theorem.

To prove the second part, we apply the Berry-Esseen Theorem [c.f. Loève (1963) p. 288] which establishes the existence of a finite constant $c$ such that

$$|H^*_n(x) - \Phi(x)| \leq c \sum_{i=2}^{n} E |Z_{ni}|^3$$

(4.3)

for all $n$ and all $x$, where $H^*_n$ is the cdf of $(T_n - \mu_n)/s_n$ and $\Phi$ is the standard normal cdf. Since $\zeta(F_n, G_n) \geq \epsilon$, we have $\sum_{i=2}^{n} E |Z_{ni}|^3 \leq 3^{3/2} \epsilon^{-3} n^{-1/2}$ and the result follows from (4.3) Q.E.D.

To compare the asymptotic performance of $T_n$ with that of $T^*_n$ and the comparable Wilcoxon rank tests, we evaluate the well-known Pitman measure of asymptotic relative efficiency (ARE). Consider the sequence of local shift alternatives

$$\{Q_n\} = \{(F_n, G_n); F_n(x) = F(x), G_n(x) = F(x - \delta n^{-1/2}), \delta > 0 \}$$

and $F \in \mathcal{F}_\delta$ (4.4)

where $\mathcal{F}_\delta$ is the class of all cdf's possessing an absolutely continuous square integrable density.

Note that under $\{Q_n\}$, $\zeta^2(F_n, G_n) \to 1/12$ as $n \to \infty$ and that

$$\lim_{n \to \infty} n^{1/2} \frac{\mu_n - 1}{(n^2 - 1/4)} = \lim_{n \to \infty} (n^{1/2}/2) \int [F(x + \delta n^{-1/2}) - F(x)] dF(x)$$

$$= \delta \pi/2$$

where
\[ \eta = \int_{-\infty}^{\infty} f^2(x) \, dx. \quad (4.5) \]

It then follows directly from Theorem 4.1 that

\[ \lim_{n \to \infty} \mathcal{L}(6n^{1/2} T_n^* \left[ \frac{n}{n^2} - \frac{1}{4} \right] |Q_n) = \mathcal{N}(3\delta \eta, 1). \quad (4.6) \]

The same method of proof as that of Theorem 4.1 establishes the limit

\[ \lim_{n \to \infty} \mathcal{L}(48n^{1/2} T_n^* \left[ \frac{n}{n^2} - \frac{1}{4} \right] |Q_n) = \mathcal{N}((12)^{1/2} \delta \eta, 1). \quad (4.7) \]

Let \( \frac{W_n}{n_1, n_2} \) denote the Wilcoxon rank-sum test statistic (that is, the sum of the treatment ranks in the combined sample) where \( n_1 \) and \( n_2 \) are the sample sizes from the control and the treatment respectively and \( n_1 + n_2 = n \). Further let \( W_n^* \) be the Wilcoxon statistic with the sample size \( \binom{n-1}{2} \) for the control and \( n \) for the treatment. A unified treatment of the asymptotic distribution theory of two sample rank order tests under local shift alternatives has been given by Hájek and Sidák (1967) using the principle of contiguity of probability measures. The next result follows directly from their Theorem VI 2.3, p. 215, if we make the identification \( \Delta = \delta n^{-1/2} \), \( \varphi(u) = u \), \( \varphi(u, f_0) = -f'(F^{-1}(u))/f(F^{-1}(u)) \) and note that

\[ \int_{0}^{1} \varphi(u) \varphi(u, f_0) \, du = \int_{-\infty}^{\infty} f^2(x) \, dx = \eta. \]

We therefore omit the proof.

**Theorem 4.2.** Assume that \( F(x) \) has an absolutely continuous density \( f(x) \) and that \( \int_{-\infty}^{\infty} \left[ f'(x)/f(x) \right]^2 f(x) \, dx < \infty. \) Then

\[ \lim_{n \to \infty} \mathcal{L}(\frac{12}{n} \left[ \frac{2W_n^*}{n(n+1)} - \frac{n}{2} \right] |Q_n) = \mathcal{N}(\delta \eta(12)^{1/2}, 1). \quad (4.8) \]
Also, if \((n_1/n) \to \lambda\) as \(n \to \infty\) and \(0 < \lambda < 1\)

\[
\lim_{n \to \infty} \mathcal{L}(W_{n_1, n_2}^{n_1/n_2, n_2/n_2} Q_n) = \mathcal{N}(\delta \eta [12 \lambda (1-\lambda)]^{1/2}, 1).
\] (4.9)

Let the symbol \(e(A:B)\) denote the Pitman ARE of a test \(A\) with respect to a test \(B\). Employing the results (4.6)-(4.9) and applying Noether's theorem (1955) on the evaluation of the Pitman ARE, we obtain the following:

\[
e_1 = e(T:T) = 3 [4 \lambda (1-\lambda)]^{-1} \geq 3,
\]

\[
e_2 = e(T:T^*) = 3/4 = e(T:T^*),
\] (4.10)

\[
e_3 = e(T^*:W^*) = 1.
\]

These are discussed in the next section.
5. COMPARISONS AND COMMENTS

The purpose of the present section is to compare the performances of the Cronholm-Revusky test $T$, the modified test $T^*$ and the two Wilcoxon tests $W$ and $W^*$ in the light of the previous results. The subscripts are suppressed throughout this section. One should bear in mind that these tests are being considered for the problem of testing for a difference between a treatment and a control in the following situations which make the Cronholm-Revusky sampling scheme relevant.

(i) The treatment has an irreversible effect on a unit while the effect of the control is purely temporary.

(ii) The cost of the experiment is primarily determined by the number of units used rather than by the total number of replications of the treatment and control.

If the data is of type (b) (c.f. Section 1), comparisons must only be drawn between $T$ and $W$ because the other two tests are not applicable. Expression (4.10) for $e_1$ shows that $T$ is three times as efficient as $W$ for shifts in any cdf $F$. This is, however, only a large sample measure of local power.

For small sample comparisons, the exact power of the Wilcoxon test for normal shift alternatives and the Lehmann alternatives may be obtained from Milton (1966) and Shorack (1967) respectively. Matching entries from their tables with the corresponding entries in our Table 1 and 2, one readily observes that the power of $T$ is considerably higher than that of $W$ for the cases considered. The deficiency in $W$ results primarily from the fact that it is based on substantially fewer replications than are included in $T$.

The criticism should therefore be leveled against the sampling scheme
involved in the Wilcoxon test \( W \) rather than against the test itself. The above comparison shows that in the situation (b), the single stage rank sum test can be greatly improved by carrying out independent subexperiments as introduced by Cronholm and Revusky (1965). Incidentally, this also substantiates Cronholm and Revusky's claim of superiority of \( T \) on the basis of their "Sensitivity Index."

When scores are available as in the case (a), a Wilcoxon test may be constructed using as many replications as are involved in \( T \) and this is the test \( W^* \). Consideration of \( W \) in this case is pointless. The competitors are, in fact, \( T \), \( T^* \) and \( W^* \). The distribution theory of the first two is substantially simpler than that of \( W^* \) since these are sums of independent components. For all parent distributions, the ARE of \( T \) relative to \( W^* \) is 75\% and that of \( T^* \) relative to \( W^* \) is 100\%. Thus the modified test \( T^* \) combines the advantages of an easy distribution theory as that of \( T \) with higher efficiency as that of \( W^* \). The exact powers of \( T \) and \( T^* \) for normal shift alternatives and the sample sizes \( n = 5, 7, 9 \) and 11, are compared in Table 4. Their nonrandomized significance levels are different. For effective comparison, we therefore use randomization in one of them (\( T^* \)) to match with the nonrandomized level of the other (\( T \)). Note that the power of \( T^* \) is uniformly higher.

A possible violation of the assumption of identical distributions for the control responses \( X_{ij} \) could occur because the units in the \( i^{th} \) stage already have received the control \((i-1)\) times, \( i = 1, 2, \ldots, n \). For instance, the history of a unit at the fifth stage is different from that at the first with respect to experiencing the control besides the possible effect of a time lag. If these are likely to be contributing factors to the response, the different
Table 4. Power of $T_n^*$ and $T_n$ for Normal shift alternatives.

(Entries for $T_n$ within braces)

<table>
<thead>
<tr>
<th>n</th>
<th>0</th>
<th>.4</th>
<th>.8</th>
<th>1.5</th>
<th>2.0</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.0083</td>
<td>.0324</td>
<td>.0935</td>
<td>.3262</td>
<td>.5499</td>
<td>.8790</td>
</tr>
<tr>
<td></td>
<td>(0.0306)</td>
<td>(0.0858)</td>
<td>(0.3002)</td>
<td>(0.5171)</td>
<td>(0.8610)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0417</td>
<td>.1244</td>
<td>.2775</td>
<td>.6379</td>
<td>.8386</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1172)</td>
<td>(0.2553)</td>
<td>(0.5935)</td>
<td>(0.8004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.0151</td>
<td>.0802</td>
<td>.2555</td>
<td>.7255</td>
<td>.9263</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0703)</td>
<td>(0.2135)</td>
<td>(0.6358)</td>
<td>(0.8673)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0681</td>
<td>.2390</td>
<td>.5236</td>
<td>.9143</td>
<td>.9883</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2162)</td>
<td>(0.4636)</td>
<td>(0.8603)</td>
<td>(0.9697)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>.0124</td>
<td>.0941</td>
<td>.3442</td>
<td>.8762</td>
<td>.9870</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0787)</td>
<td>(0.2746)</td>
<td>(0.7802)</td>
<td>(0.9549)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0597</td>
<td>.2701</td>
<td>.6261</td>
<td>.9726</td>
<td>.9988</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2367)</td>
<td>(0.5439)</td>
<td>(0.9349)</td>
<td>(0.9932)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>.0083</td>
<td>.0917</td>
<td>.3895</td>
<td>.9345</td>
<td>.9969</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0741)</td>
<td>(0.3042)</td>
<td>(0.8555)</td>
<td>(0.9830)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0433</td>
<td>.2632</td>
<td>.6684</td>
<td>.9884</td>
<td>.9998</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2249)</td>
<td>(0.5747)</td>
<td>(0.9632)</td>
<td>(0.9980)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance levels are nonrandomized for $T_n$ and randomized for $T_n^*$.

Subexperiments are to be treated as blocks. In the presence of such block effects, one can easily verify that the two sided version of the test $T$ coincides with a rank test proposed by Wilcoxon (1946) [see also Hodges and Lehmann (1962)]. The tests $T^*$ and $W^*$ are not applicable in this case.

Finally, we remark that unlike the Wilcoxon test, $T$ and its modification $T^*$ are not locally most powerful for the logistic shift alternatives.
However, they do perform quite well and the distribution theory is simple. Many modifications of the Cronholm-Revusky sampling scheme are also possible and these will be discussed elsewhere.

6. ACKNOWLEDGEMENT

The authors would like to take this opportunity to thank Professor Henry Neave, currently at the University of Nottingham, for his assistance in programming the computations. This research was supported by the Office of Naval Research under contract Nonr 1202(17), Project No. 042-222 and by the Wisconsin Alumni Research Foundation through grants 67-08500-1 and 67-45435-1 for computer usage.
REFERENCES


