ASYMPTOTIC EFFICIENCY OF TWO NONPARAMETRIC COMPETITORS
OF WILCOXON'S TWO SAMPLE TEST
by
MYLES HOLLANDER
FSU Statistics Report M109

(Revision of FSU Statistics Report M109 entitled 'Wilcoxon's
Signed Rank Test As A Large Sample Competitor of Wilcoxon's
Rank Sum Test.)

Research Supported
by
National Institutes of Health
Under Grant GRNT 17-946

July, 1966
The Florida State University
Tallahassee, Florida
Asymptotic Efficiency of Two Nonparametric Competitors
of Wilcoxon's Two Sample Test
by Hyles Hollander
The Florida State University

1. Introduction

Let \( X_1, \ldots, X_m \) be independent and identically distributed according to \( F_1 \) and \( Y_1, \ldots, Y_n \) be independent and identically distributed according to \( F_2 \) where \( F_1, F_2 \) are assumed continuous. The excellent properties of the Mann-Whitney-Wilcoxon U statistic ([12], [7]) for testing \( H_0: F_1 = F_2 = F(\text{unknown}) \) against translation alternatives are well known (e.g. see [3], [4]) and may be attributed, in part, to the fact that \( U/mn \) is the uniform minimum variance unbiased estimator of \( P(X_1 < Y_1) \). This suggests the investigation of tests based on statistics which (when suitably normed) are consistent estimators of \( P(X_1 + X_2 < Y_1 + Y_2) \). In this paper we concern ourselves with two statistics having this property.

The \( W \) statistic, to be used only when \( m=n \), is defined to be Wilcoxon's signed rank statistic [12] applied to a random pairing of the \( X \)'s with the \( Y \)'s. The second statistic \( V^* \), which is the uniform minimum variance unbiased estimator of \( P(X_1 + X_2 < Y_1 + Y_2) \), is the proportion of the \( \binom{m}{2} \binom{n}{2} \) quadruples \((X_i, X_j; Y_k, Y_\ell)\) with \( i < j \) and \( k < \ell \) satisfying the inequality \( X_i + X_j < Y_k + Y_\ell \).

In section 2 we show that \( V \) is not distribution-free under \( H_0 \) but we define an asymptotically distribution-free procedure based on asymptotic normality and a consistent estimate of the null asymptotic variance of \( V \). Section 3 is devoted to efficiency comparisons of \( U, V, \) and \( W \) for translation.

\footnote{The \( V \) statistic was suggested for consideration by Professor John Pratt in the review of an earlier version of this paper.}
and contamination alternatives. In particular we find that \( V \) and \( W \) can be more Pitman efficient than \( U \) for contamination with a shift. Section 4 contains the author's recommendations for the use of these procedures which, roughly speaking, are i) when \( m=n \) and \( n \) is not too small (say \( n \geq 10 \)) prefer \( W \) to \( U \) and \( V \), ii) for \( m \neq n \) prefer \( U \) to \( V \).

2. Definitions and Basic Facts.

The Mann-Whitney form of Wilcoxon's statistic is

\[
U = \sum_{i=1}^{m} \sum_{j=1}^{n} \phi(X_i, Y_j)
\]

where \( \phi(a,b) = 1 \) if \( a < b \); 0 otherwise. To define the \( W \) statistic, let us assume for simplicity (and without loss of generality) that the random pairing of the \( X \)'s with the \( Y \)'s results in the pairs \( (X_i, Y_i) \), \( i=1,\ldots,n \). Let \( D_i = |X_i - Y_i| \) and \( R_i = \text{rank of } D_i \) in the joint ranking from least to greatest of \( [D_i]_i \). Then Wilcoxon's signed rank statistic is

\[
W = \sum_{i=1}^{n} R_i \phi(X_i, Y_i).
\]

Using a representation due to Tukey [11], we may write \( W \) as

\[
W = \sum_{i<j}^{n} \psi(X_i, X_j; Y_i, Y_j) + \sum_{i=1}^{n} \psi(X_i, X_i; Y_i, Y_i),
\]

where \( \psi(a,b;c,d) = 1 \) if \( a+b < c+d \); 0 otherwise. Letting \( W' \) denote the first term on the right of (2.3), we note that \( W \) and \( W' \) are asymptotically equivalent test statistics and that \( (2W'/n(n-1)) \) is an unbiased and consistent estimator of
P(X_1+X_2 < Y_1+Y_2) as is V, defined as

\[ V = \left( \begin{array}{c} m \\ 2 \end{array} \right)^{-1} \left( \begin{array}{c} n \\ 2 \end{array} \right)^{-1} \sum_{i<j}^{\infty} \psi(X_i, X_j; Y_k, Y_{\ell}). \] (2.4)

We remark that the statistic V, even when m=n, is based on more "information" than W' as the indicator function \( \psi \) is computed for \( n^2 (n-1)/4 \) quadruples in the case of V versus \( n(n-1)/2 \) for W'. Furthermore, unlike U and W, V is not distribution-free under \( H_0 \). From Lehmann's generalized U-statistic theorem [5] we may immediately state

**Theorem 1:** If \( 0 < \int F_1 dF_2 < 1 \), and \( \lim(m/n) = c \), \( (0 < c < 1) \), then

\( (m)^{3/2} (V - P(X_1+X_2 < Y_1+Y_2)) \) has a limiting normal distribution.

Under the null hypothesis, the asymptotic variance of V may be written in the form

\[ \sigma^2_A(V) = \left( \begin{array}{c} m \\ 2 \end{array} \right)^{-1} \left( \begin{array}{c} n \\ 2 \end{array} \right)^{-1} \left[ 2(n-2)\left( \begin{array}{c} m-2 \\ 2 \end{array} \right) + 2(m-2)\left( \begin{array}{c} n-2 \\ 2 \end{array} \right) \right] \lambda(F) - (1/4) \] (2.5)

where

\[ \lambda(F) = P(X_1 < X_2 + X_3 - X_4; X_1 < X_5 + X_6 - X_7) \] (2.6)

when \( X_1, X_2, \ldots, X_7 \) are independent and identically distributed according to F.

Lehmann [6] has obtained different values of \( \lambda(F) \) for various F and thus the null distribution of V will depend on F. Hence, in the remainder of this paper the phrase "the V test" will mean the asymptotically distribution-free procedure which treats \( (V - (1/2))/\sigma_A(V) \) as a unit normal random variable under \( H_0 \) where \( \hat{\sigma}_A^2(V) \) is defined by replacing \( \lambda(F) \) with a consistent estimate in (2.5).

One such estimate, proposed by Lehmann [6] in another context, is the following.
Let \( Z_1, Z_2, \ldots, Z_{m+n} \) denote the combined sequence of \( X \)'s and \( Y \)'s and define \( \hat{\lambda}(F) \) to be the relative frequency of the event \( (Z_{\alpha_1} < Z_{\alpha_2} + z_{\alpha_3} - z_{\alpha_4}; Z_{\alpha_5} < Z_{\alpha_6} + z_{\alpha_7}) \) over a small subset of the total number of such simultaneous inequalities that could be checked.

3. Efficiencies for Translation and Contamination Alternatives

We first compare, on the basis of Bahadur efficiency ([1], [2]), the asymptotic performance of the \( U, V, \) and \( W \) tests for the translation alternatives \( F_1(x) = F(x), F_2(x) = F(x-\theta) \). Let \( T_1, T_2 \) denote any two of the statistics \( U, V, W \), let \( N = m+n \), and define \( c_1(\theta) = p_{\theta} \lim \left[ (T_1 - E_0(T_1)) / \sqrt{\sigma_0(T_1)} \right] \). Then from the asymptotic normality of \( U \) and \( V \) and \( W \) it follows that the Bahadur efficiency \( B_\theta(T_1, T_2) \) of the \( T_1 \) test with respect to the \( T_2 \) test is given by \( \left[ c_1(\theta) / c_2(\theta) \right]^2 \).

The probability limits for each test are easily calculated and we thus state

Theorem 2: For the translation alternatives \( F_1(x) = F(x), F_2(x) = F(x-\theta) \),

\[
B_\theta(U, U) = \frac{\left( [G(x+2\theta)dG(x)-(1/2)]^2 \right)}{2(\int [F(x+\theta)dF(x)-(1/2)]^2)}, \tag{3.1}
\]

\[
B_\theta(V, U) = \frac{\left( [G(x+2\theta)dG(x)-(1/2)]^2 \right)}{48(\lambda(F)-(1/4)) \int [F(x+\theta)dF(x)-(1/2)]^2}, \tag{3.2}
\]

\[
B_\theta(V, W) = (24\lambda(F)-6)^{-1}. \tag{3.3}
\]

where \( G \) is the distribution function of \( X_1 - X_2 \) when \( X_1, X_2 \) are independent and identically distributed according to \( F \) and \( \lambda(F) \) is given by (2.6).
The following table gives values of the Bahadur efficiencies when $F$ is normal.

<table>
<thead>
<tr>
<th>$\theta/\sigma$</th>
<th>0.25</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_\theta(W,U)$</td>
<td>0.990</td>
<td>0.960</td>
<td>0.860</td>
<td>0.641</td>
<td>0.533</td>
</tr>
<tr>
<td>$B_\theta(V,U)$</td>
<td>1.03</td>
<td>0.995</td>
<td>0.891</td>
<td>0.665</td>
<td>0.552</td>
</tr>
<tr>
<td>$B_\theta(V,W)$</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
</tr>
</tbody>
</table>

We note that $B_\theta(V,W)$ is independent of $\theta$ and letting $\theta$ tend to $\infty$ in (3.1) and (3.2) yields

Corollary 2.1: $\lim_{\theta \to \infty} B_\theta(W,U) = 0.5$. $\lim_{\theta \to \infty} B_\theta(V,U) = (48\lambda(F)-12)^{-1}$.

Lehmann [6] has shown that $\lambda(F) \leq 7/24$ which implies

Corollary 2.2: For all $F$, $B_\theta(V,W) \geq 1$ and $\lim_{\theta \to \infty} B_\theta(V,U) \geq 0.5$.

Furthermore, by letting $\theta$ tend to 0 in (3.1)-(3.3) and applying a result of Bahadur [1] we may state

Corollary 2.3: The Pitman efficiencies for the alternatives $F_2^{(n)}(x) = F(x-(c/\sqrt{n}))$ are

$$E(W,U) = 2(\frac{g^2}{f^2})^2,$$  \hspace{1cm} (3.4)

$$E(V,U) = (12\lambda(F)-3)^{-1}(\frac{g^2}{f^2})^2,$$  \hspace{1cm} (3.5)

$$E(V,W) = (24\lambda(F)-6)^{-1}.$$  \hspace{1cm} (3.6)

where $f$ and $g$ are the densities corresponding to $F$ and $G$, respectively.

Equation (3.4) should not be regarded as new as it is implicit in the work of Pitman [9] where the efficiencies of both the signed rank and rank sum tests are given.
Some values of the Pitman efficiencies are:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( E(W, U) )</th>
<th>( E(V, U) )</th>
<th>( E(V, W) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( f(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} ), ( -\infty &lt; x &lt; \infty )</td>
<td>1.00</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>2. ( f(x) = 1, 0 \leq x \leq 1; 0 ) otherwise.</td>
<td>.889</td>
<td>.906</td>
<td>1.02</td>
</tr>
<tr>
<td>3. ( f(x) = e^{-x}, x &gt; 0; 0 ) otherwise.</td>
<td>.500</td>
<td>.529</td>
<td>1.06</td>
</tr>
</tbody>
</table>

We note that the V test is more Pitman efficient than the U test when F is normal but, in general, the values in Table 3.2 slightly favor U over both V and W. However, we now turn to a situation where V and W are more (Pitman) efficient than U.

We consider the contamination alternatives

\[ F_1(x) = F(x), F_2(x) = (1-p)F(x) + pH(x), \] 

with \( H \leq F \) and compare the Pitman efficiency of the U, V, W tests for the hypothesis \( p = 0 \). (Hodges and Lehmann [3] have used these alternatives for comparing the U test with the normal theory t-test.) A straightforward calculation shows

\[
E_p \left( \frac{2W'}{n(n-1)} \right) = E_p(V) = \int (F_1 \ast F_1) d(F_2 \ast F_2) \\
= 1 - \int [(1-p)^2 F \ast F + 2p(1-p) F \ast H + p^2 H \ast H] d(F \ast F)
\]

and thus

\[
\frac{d}{dp} E_p \left( \frac{2W'}{n(n-1)} \right) \bigg|_{p=0} = \frac{d}{dp} E_p(V) \bigg|_{p=0} = 1 - 2 \int (F \ast H) d(F \ast F)
\]

(3.7)

where "\( \ast \)" denotes convolution. Hodges and Lehmann show

\[
\frac{d}{dp} E_p(U/mn) \bigg|_{p=0} = \int F dH - 1/2
\]

(3.8)
and a direct application of Pitman's formula [8] then yields

**Theorem 3:** For the contamination alternatives

\[ F_1(x) = F(x), F_2(x) = (1-p)F(x) + pH(x), \] 

the Pitman efficiencies \((p \rightarrow 0)\) are

\[
E(U, U) = \frac{2[(1/2)-\int (F \ast H) d(F \ast F)]^2}{[\int F dH-(1/2)]^2}, \tag{3.9}
\]

\[
E(V, U) = \frac{[(1/2)-\int (F \ast H) d(F \ast F)]^2}{(12\lambda(F)-3)[\int F dH-(1/2)]^2}, \tag{3.10}
\]

\[
E(V, W) = (24\lambda(F)-6)^{-1}. \tag{3.11}
\]

Table 3.3 contains values of the Pitman efficiencies when \(H(x) = F(x-\theta)\) and \(F\) is normal.

<table>
<thead>
<tr>
<th>(\theta/\sigma)</th>
<th>.25</th>
<th>.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E(U, U))</td>
<td>1.01</td>
<td>1.02</td>
<td>1.08</td>
<td>1.31</td>
<td>1.61</td>
</tr>
<tr>
<td>(E(V, U))</td>
<td>1.04</td>
<td>1.06</td>
<td>1.12</td>
<td>1.36</td>
<td>1.67</td>
</tr>
<tr>
<td>(E(V, W))</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
<td>1.04</td>
</tr>
</tbody>
</table>

**Corollary 3.1:** For the contamination alternatives with \(H(x) = F(x-\theta)\),

\[
\lim_{\theta \rightarrow \infty} E(W, U) = 2, \quad \lim_{\theta \rightarrow \infty} E(V, U) = (12\lambda(F)-3)^{-1}.
\]

**Corollary 3.2:** For all \(F\), \(\lim_{\theta \rightarrow \infty} E(V, U) \geq 2.\)
4. Conclusions and Recommendations

For the equal sample size case, with n moderately large, the author prefers the random-paired signed rank test \( U \) to the Wilcoxon \( U \) test since i) \( W \) can be more Pitman efficient than \( U \) for contamination with a shift, ii) often there is little loss in Pitman efficiency when using \( W \) in place of \( U \) (there is no loss at all when \( F \) is normal), iii) the \( W \) test will maintain its significance level when \( F_1 \) and \( F_2 \) differ only by a scale parameter while the \( U \) test will not be exact in such cases (e.g., see [10]), iv) I find it easier to rank the \( n \) absolute differences than the original combined sample of \( 2n \) observations, and v) I believe the above properties overshadow the disadvantage of basing a decision on a test which utilizes an irrelevant random mechanism.

In the choice between \( W \) and \( V \) when \( m=\bar{n} \), \( W \) is again recommended since ii) \( V \) is tedious to compute and is not exactly distribution-free under \( H_0 \), and ii) the efficiency of \( W \) with respect to \( V \), given by the expression \( (24\lambda(F)-6) \), is very close to one. The values for \( F \) normal, rectangular, and exponential are .965, .981, and .944, respectively.

When \( m\neq n \), the author prefers the \( U \) test to the \( V \) test despite the fact that the Pitman efficiency of \( V \) with respect to \( U \) is slightly greater than one for normal translation. This preference is based on the fact that \( U \) is distribution-free under \( H_0 \) and considerably easier to compute than \( V \). However, if the experimenter has reason to expect contamination, then the efficiency calculations of section 3 give support to the choice of \( V \) over \( U \).

As a final remark we mention that the set of alternatives for which the \( U \) test is consistent is different than the set of alternatives for which the \( V \) and \( W \) tests are consistent. The two-sided test based on \( U \) is consistent
if and only if $P(X_1 < Y_1) \neq 1/2$ while the two-sided tests based on $V$ and $W$ are consistent if and only if $P(X_1 + X_2 < Y_1 + Y_2) \neq 1/2$. If we consider the densities $f_1(x) = 1$ if $4 \leq x \leq 5$, and 0 otherwise, and $f_2(x) = a$ if $1 \leq x \leq 2$, $b$ if $10 \leq x \leq 11$, and 0 otherwise, a simple calculation shows that for $a = b = 1/2$ the two-sided $W$ and $V$ tests are consistent but the two-sided $U$ test is not consistent, whereas for $a = 1/\sqrt{2}$, $b = 1-(1/\sqrt{2})$ we get the opposite conclusion. Similar examples are easily constructed for the one-sided tests.

5. Acknowledgment

I am grateful to the late Professor Frank Wilcoxon for suggesting the comparison between the signed rank and rank sum tests in the equal sample size case and for many interesting and valuable discussions relating to these procedures.
REFERENCES


