ON THE LIMITING PITMAN EFFICIENCY OF SOME
RANK TESTS OF INDEPENDENCE

BY

TERESA LEDVINA

TECHNICAL REPORT NO. 202
OCTOBER 1984

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
2FFZ459
INGRAM OLKIN, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE LIMITING PITMAN EFFICIENCY OF SOME
RANK TESTS OF INDEPENDENCE

BY

Teresa Ledwina

TECHNICAL REPORT NO. 202
October 1984

Prepared Under the Auspices
of
National Science Foundation Grant MCS81-04262
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE LIMITING PITMAN EFFICIENCY OF SOME RANK TESTS
OF INDEPENDENCE

Teresa Ledwina

Institute of Mathematics, Technical University of
Wrocław, Wrocław, Poland
The limiting (as the significance level approaches 0) Pitman efficiency of a new "regression-based" rank test of independence to Kendall's tau and Spearman's rho is derived. The result is based on a theorem of Wieand (Ann. Statist. 4 (1976), 1003-1011) on coincidence of the limiting Pitman efficiency and the local (as the alternative approaches the hypothesis) approximate Bahadur efficiency.

For each pair of the considered tests, the equality of the limiting Pitman efficiency, local approximate and exact local Bahadur efficiency is also shown. The latter fact is strongly related to some recent results of Kremer (J. Multivar. Anal. 11 (1981), 532-534) on the local equality of exact and approximate Bahadur efficiency of some rank tests of independence.


Key words and phrases: Asymptotic theory, Bahadur efficiency, Pitman efficiency, rank tests of independence.
1. INTRODUCTION AND SUMMARY

There are several ways to introduce nonparametric test statistics of bivariate independence. We mention here two intuitively appealing. The first one is to consider some estimators of measures of dependence. For example, Kendall's tau, Spearman's rho and Blomqvist's quadrant tests can be introduced in this way. These tests are usually suitable to detect positive (or negative) dependence. The second way is to use some distance measures between the bivariate empirical distribution function and the product of marginals, which leads to a generalization of the Kolmogorov-Smirnov and Cramér-von Mises statistics (cf. Blum et al. [6]). The former tests have excellent power properties, but the null distribution of the first one is not known. The second test statistic is a little easier to investigate and some properties of it has been proved by Blum et al. [6], Neuhaus [26] and De Wet [9], among others. However, contrary to the first group of tests, Cramér-von Mises test for independence is destined to detect arbitrary deviations from independence.

In Ledwina [21] a new test statistic for testing independence against positive (or negative) dependence has been introduced. From one hand, this statistic is based on an estimator of a measure of dependence introduced in Kowalczyk and Ledwina [15] as a version of
monotonic dependence function considered earlier by Kowalszyk and Pleszczyńska [16]. The measure takes values in \( C[0,1] \) and can be considered as a regression-based measure of dependence (see Dąbrowska [8]). On the other hand, the statistic is the supremum over a "time interval" \((0,1)\) of a process which behaves asymptotically as the usual empirical process on which the original Kolmogorov–Smirnov test is based (cf. Bednarski and Ledwina [2]).

To define the new statistic, let us denote by \((X_i, Y_i), \ldots, (X_n, Y_n)\) a set of independent identically distributed random vectors and let \(R_i(S_i)\) be the rank of \(X_i(Y_i)\) among \(X_1, \ldots, X_n(Y_1, \ldots, Y_n)\). Moreover, let

\[
T_n(p) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{R_i}{n} \right) \left( p - I(S_i \leq np) \right), \quad p \in (0,1),
\]

where \(I(A)\) is the indicator of the set \(A\). Then the test statistic is of the form

\[
t_n = \sup_{0 < p < 1} (12n)^{1/2} T_n(p).
\]

In this paper, the limiting (as the significance level approaches 0) Pitman efficiency of \(t_n\) to \(\tau\) and \(\rho\) is derived. The result is based on a version of a theorem of Wieand [30], which gives conditions under which the existence of the limiting (as the alternative approaches the hypothesis) approximate Bahadur efficiency implies the existence of the limiting Pitman efficiency and the equality of the two limits. It should be said that Wieand’s result is applicable for large class of linear rank tests (see Section 3.2).
However, to avoid introducing some additional notation and quoting some well known results we restrict attention to the Spearman's test - a member of this class.

In Ledwina[21], the exact Bahadur slope of $t_n$ was derived and an expansion of it was given. Hence, local coincidence of the approximate and exact Bahadur slopes of $t_n$ easily follows. The same holds true for large class of linear rank tests of independence as well as for Kendall's tau (cf. Kremer[19]). Combining these results with those derived in this paper we conclude that the limiting Pitman efficiency and local Bahadur efficiencies are equivalent for large class of rank tests of independence. In this context, let us mention that in many recent papers on linear rank tests (see Behnken[3,4], Kremer[18,20], Ruymgaart[28]) so called asymptotic relative efficiency (ARE) of tests has been derived. Often the ARE is called Pitman efficiency, also. The ARE, introduced by Hájek and Šidák[12], is based on comparison of asymptotic powers of tests. As exemplified by Kremer[18,20], local coincidence of the ARE, approximate and exact Bahadur efficiency holds for some linear rank tests, also. Still, since $t_n$ behaves like Kolmogorov-Smirnov statistic, to compare $t_n$ with some asymptotically normally distributed rank tests the Wieand's approach to Pitman efficiency is more suitable.

The organization of the paper is as follows. In Section 2 we restate in a suitable form assumptions of the Wieand [30] theorem. Some of the assumptions are checked in Section 2, also. The crucial assumption (Condition III*) is verified in Section 3. Some examples of the limiting Pitman efficiency are given in Section 4.
2. PRELIMINARIES

Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be a set of independent identically distributed random vectors, with common continuous bivariate distribution function \(H(x, y)\) having marginal df’s \(F(x)\) and \(G(y)\).

We are concerned with testing the null hypothesis

\[ H = F \cdot G \]

against the positive quadrant dependence (pqd) of \(X\) and \(Y\), i.e.

\[ H \not\prec F \cdot G. \]

The test statistics we consider are distribution free under the null hypothesis and the probability integral transformation implies that we can carry out all calculations for so-called uniform representation of \(H\). Consequently, we assume throughout \(H(u, v)\) has both marginals uniform on \((0, 1)\).

To derive the approximate Bahadur efficiencies of the considered tests we shall prove first of all that the corresponding sequences of test statistics are standard sequences in the sense of Bahadur[1].

Recall that a sequence \(\{s_n\}\) of real-valued statistics is said to be a standard sequence for the above testing problem if there exist

I. A continuous df \(D\) such that under the null hypothesis

\[ \lim_{n \to \infty} P(s_n < x) = D(x) \quad \text{for each } x. \]
II. A constant \( a, \ 0 < a < \infty \), such that \( \log(1-D(x)) = -(ax^2/2)(1+o(1)) \) where \( o(1) \to 0 \) as \( x \to \infty \).

III. A real-valued function \( b(H) \) such that for every \( H \) from the alternative \( 0 < b(H) < \infty \) and 
\[
\lim_{n \to \infty} P_H(|s_n^{1/2} - b(H)| > x) = 0 \quad \text{for every} \quad x > 0 .
\]

REMARK 2.1. Kendall's tau and Spearman's rho form standard sequences with \( a=1 \) and

\[
b(H) = E(H,K) = 6 \sum_{00}^{11} H(u,v) dH(u,v) - 3/2 ,
\]

\[
b(H) = E(H,S) = 12 \sum_{00}^{11} uv dH(u,v) - 3 ,
\]
respectively.

REMARK 2.2. \( \{t_n\} \) is the standard sequence.

PROOF. By Bednarski and Ledwina[2], under independence of
\( X \) and \( Y \),

\[
(12n)^{1/2} T_n \to W_0 \quad \text{in} \quad D[0,1] ,
\]

where \( W_0 \) is the Brownian bridge and \( T_n(0) = T_n(1) = 0 \).

Hence, Conditions I and II hold with \( D(x) = 1-\exp(-2x^2) \),
\( x > 0 \), and \( a=4 \). Moreover, by Lemma 3.1 of Ledwina[21],
\( \{t_n\} \) satisfies III with
b(H) = B(H,T) = \( (12)^{1/2} \sup_{0<p<1} B(H,p) \),

where

\[
B(H,p) = \int_{00}^{11} u(p - I(v \leq p))dH(u,v).
\]

As shown by Wieand [30] a stronger version of III is needed to derive the limiting Pitman efficiency via the local approximate Bahadur efficiency. The main result of Wieand [30], p. 1005, is stated and proved in the case when the distribution of the observed random variables depends on a real parameter \( \theta \) which ranges over an interval and the null hypothesis is a single point. However, as easily seen, the result can be stated for more general situations. The proof requires only obvious changes. A possibility of such extension was previously noticed by Kremer [17].

To specify the above to the considered problem set

\[ H_0(u,v) = uv \text{ for } (u,v) \in (0,1)^2 \]

and

\[ d(H_1,H_2) = \sup_{0 \leq u,v \leq 1} |H_1(u,v) - H_2(u,v)| \text{ for any df's } H_1,H_2. \]

Then, the main assumption of the Wieand paper (Condition III*) can be replaced by

III*. Suppose for a standard sequence \( \{ s_n \} \) there is a \( d^* > 0 \) such that for every \( \varepsilon > 0 \) and \( \delta \in (0,1) \), there is a \( C \) such that for all \( H \in \{ H : d(H,H_0) < d^* \} \), we have for all \( p,q,d \), and \( n > C/b^2(H) \) we have
\[ P_H \left\{ \left| n^{-1/2} s_n - b(H) \right| < \varepsilon b(H) \right\} > 1 - \delta. \]

Note that in the above definition \( C \) may depend on \( d^* \) but is otherwise independent of \( H \).

3. VERIFICATION OF CONDITION III*

Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be i.i.d. random vectors defined on a measurable space \((S, \mathcal{S})\) each distributed according to a continuous df \( H(u, v) \) with uniform marginals on \((0, 1)\). Denote by \( H_n(x, y) \) the right-continuous empirical df and by \( F_n(x) \) and \( G_n(y) \) the marginals of \( H_n(x, y) \).

To verify III* for \( \tau \) and \( t_n \), the following version of Theorem 1 of Kiefer and Wolfowitz[13] will be used.

THEOREM 3.1. There exist positive constants \( c_0 \) and \( c \) such that, for all \( n \), all \( H \), and all positive \( r \),

\[ P_H \left( \sup_{0 \leq u, v \leq 1} \left| H_n(u, v) - H(u, v) \right| < n^{-1/2} r \right) > 1 - c_0 \exp(-cr^2). \]

We shall apply also the following result (quoted in Marshall-Olkin[24], p.36) on integration by parts in two dimensions.

LEMMA 3.1. If \( A(0, y) = A(x, 0) = 0 \), \( A \) is of bounded variation on finite intervals and \( J(x, y) = P(X \leq x, Y \leq y) \) then
\[
\sum_{0}^{\infty} A(x,y)dJ(x,y) = \sum_{0}^{\infty} P(X>x, Y>y)dA(x,y).
\]

Recall also that \( B(H,K) = \frac{3\tau}{2}, \ B(H,S) = \varsigma_S, \) where \( \tau \) and \( \varsigma_S \) are Kendall's and Spearman's measures of association, respectively. By definition of the measures and Lemma 3 of Lehmann[23], we have under pqd alternatives

\[ 0 < B(H,K) \leq \frac{3}{2}, \quad 0 < B(H,S) \leq 1. \]

Besides, by properties of \( B(H,T) \) (cf. Ledwina[21]), for all \( H \) from the alternative it holds

\[ 0 < B(H,T) \leq (12)^{1/2}/8. \]

By the above, to show III* for tau and rho it is enough to prove

**III**. For every \( \delta \in (0,1) \), there is a \( L \) such that for all \( H \) and \( n > 1 \) it holds

\[ P_H( |s_n - n^{1/2} b(H)| < L ) > 1 - \delta. \]  \( (3.1) \)

3.1. KENDALL'S TAU

Kendall's tau can be expressed by
\[ K_n = \left\{ n/(n-1) \right\} \left\{ \sum_{00}^{11} H_n(u,v) dH_n(u,v) - 3/2 \right\}, \]

where

\[ n^{1/2} \left\{ K_n - B(H,K) \right\} \rightarrow N(0,1). \]

Hence, it is enough to prove III** with (3.1) replaced by

\[ P_H \left( n^{1/2} \left| \sum_{00}^{11} H_n(u,v) dH_n(u,v) - \sum_{00}^{11} H(u,v) dH(u,v) \right| < L \right) > 1 - \delta. \]

Let \( c_0 \) and \( c \) be as in Theorem 3.1. For a given \( \delta \in (0,1) \) let \( r_0 \) be the smallest \( r \) such that \( \delta > c_0 \exp(-cr^2) \). Define a subset \( S_0 \) of \( S \) by

\[ S_0 = \left\{ \sup_{04u,v \leq 1} \left| H_n(u,v) - H(u,v) \right| < r_0 n^{-1/2} \right\} . \]

By Theorem 3.1, for all \( H \) and \( n \), \( P_H(S_0) > 1 - \delta \). We shall show that for \( L = 4r_0 \), every \( H \) and \( n > 1 \) on \( S_0 \) it holds

\[ n^{1/2} \left| \sum_{00}^{11} H_n(u,v) dH_n(u,v) - \sum_{00}^{11} H(u,v) dH(u,v) \right| < L. \]

We have

\[ n^{1/2} \left| \sum H_n dH_n - \sum H dH \right| \leq \]
\[ n^{1/2} \int \int \left| H_n - H \right| dH + n^{1/2} \int \int H_n d(H_n - H). \] (3.2)

The first component of (3.2) can be majorized on \( S_0 \) by \( r_0 \). By Lemma 3.1, the second component of (3.2) equals

\[ n^{1/2} \int \int \left\{ H_n(u,v) - H(u,v) + u - F_n(u) - G_n(v) + v \right\} dH_n(u,v) \]

and is majorized on \( S_0 \) by \( 3r_0 \). Setting \( L = 4r_0 \) the proof is concluded.

3.2. SPEARMAN'S RHO

As we have already observed, to prove III* for rho it is enough to show III** - a uniform tightness condition. III** is fulfilled for rho and for a class of linear rank tests of independence by the uniform asymptotic normality of these statistics under the alternative. Some results on the uniform asymptotic normality can be found in Ruymgaart et al. [27], Ruymgaart [28] and Bönnert et al. [7], for example.

Recall also that Sinha and Wieand [29], by some straightforward calculations, have checked III* for a multivariate version of rho under an extension of Bhuchongkul [5] dependence model.

3.3. VERIFICATION OF III* FOR \( t_n \)

The condition III* for \( t_n = \sup_{0 < p < 1} (12n)^{1/2} T_n(p) \) will be checked by a similar reasoning as we have applied...
in Section 3.1. In this case, to verify III* it is enough to show

\[ \text{III**. For every } \delta \in (0,1), \text{ there is a } \Lambda \text{ such that for all } H \text{ and } n > 1 \text{ it holds} \]

\[ P_H(\sup_{0 < p < 1} |T_n(p) - \mathbb{E}(H,p)| < L) > 1 - \delta. \]

To justify the above observe that for \( n > 12L^2 \varepsilon^{-2} b^{-2}(H) \),

\[ b(H) = (12)^{1/2} \sup_{0 < p < 1} \mathbb{E}(H,p), \text{ it holds} \]

\[ P_H(\sup_{0 < p < 1} |T_n(p) - \mathbb{E}(H,p)| < \varepsilon 12^{-1/2} b(H)) \leq \]

\[ P_H(\sup_{0 < p < 1} |T_n(p) - \mathbb{E}(H,p)| < \varepsilon b(H)) \cdot \]

Now let \( S_0 \) be as in Section 3.1. We shall show that for every point of \( S_0 \) and \( L = 7r_0 \) it holds

\[ n^{1/2} \sup_{0 < p < 1} |T_n(p) - \mathbb{E}(H,p)| < L. \]

Observe that

\[ T_n(p) = \sum_{00}^{11} F_n(u)(p - I(G_n(v) \leq p))dH_n(u,v) \]

and consequently
\[ |T_n(p) - E(H, p)| \leq \]
\[ \sum_{00}^{11} (F_n(u) - u)(p - I(G_n(v) \leq p))dH_n(u, v) \] + 
\[ \sum_{00}^{11} u(p - I(G_n(v) \leq p))d(H_n(u, v) - H(u, v)) \] + 
\[ \sum_{00}^{11} u(I(v \leq p) - I(G_n(v) \leq p))dH(u, v) \] .

(3.3)

The first component of (3.3) does not exceed \( n^{-1/2}r_0 \) on \( S_0 \). To majorize the second term of (3.3) write

\[ \sum_{00}^{11} u(p - I(G_n(v) \leq p))d(H_n(u, v) - H(u, v)) = \]
\[ (p-1) \int_0^1 (u-F_n(u))du + \]
\[ \sum_{00}^{11} u(1-I(G_n(v) \leq p))d(H_n(u, v) - H(u, v)) . \]

(3.4)

The absolute value of the first term of (3.4) can be majorized on \( S_0 \) by \( n^{-1/2}r_0 \). To majorize the second component of (3.4) we shall apply Lemma 3.1 with \( A(u, v) = A_n(u, v) = u(1-I(G_n(v) \leq p)) \). Hence, the second term of (3.4) equals

\[ \sum_{00}^{11} \left\{ H_n(u, v) - H(u, v) + u-F_n(u) + v-G_n(v) \right\}dA_n(u, v) \]

and is majorized by \( 3n^{-1/2}r_0 \) on \( S_0 \).
To show that the last term of (3.3) does not exceed $2n^{-1/2}r_0$ on $S_0$, set

$$E = \{ v : I(G_n(v) \leq p) > I(v \leq p) \}.$$

On $S_0$, for every $v$, it holds $v_n^{-1/2}r_0 \leq G_n(v) \leq v + n^{-1/2}r_0$ and consequently, on $S_0$, $G_n(v) \leq p$ implies $v \leq p + n^{-1/2}r_0$.

Hence,

$$\int_E (I(G_n(v) \leq p) - I(v \leq p)) dv \leq$$

$$\int_E (I(v \leq p + n^{-1/2}r_0) - I(v \leq p)) dv \leq n^{-1/2}r_0.$$

Similar reasoning applied to the complement of $E$ completes the proof.

By the above, on $S_0$, for all $p$, all $H$ and $n$, it holds

$$|T_n(p) - E(H, p)| \leq 7n^{-1/2}r_0.$$

4. EXAMPLES

In this section some examples of the limiting Pitman efficiency of $t_n$ to tau and rho will be presented. All examples we give concern one-parameter bivariate distributions, say $H_\theta(u, v)$, where $\theta = 0$ is equivalent to independence and $pq\theta$ is equivalent to $\theta > 0$. So, the original result of Wieand[30] applies in all cases. Hence, letting $E_{T(1)T(2)}$ represent the limiting Pitman efficiency
of $T^{(1)}$ to $T^{(2)}$ we have

$$E_{T^{(1)},T^{(2)}} = \lim_{\theta \to 0} \left\{ \frac{a_1 b_1^2(\theta)}{a_2 b_2^2(\theta)} \right\}.$$  

In Ledwina[21] the exact Bahadur efficiencies were calculated in all cases considered in Sections 4.1 - 4.4, below. Moreover, the limiting exact efficiencies were derived in Ledwina[22]. Since, as was mentioned, the limiting Pitman efficiency coincides with the limiting exact Bahadur efficiency we shall quote here only some numerical values of the limiting efficiency.

4.1. LIMITING PITMAN EFFICIENCIES UNDER THE MORGENSTERN ALTERNATIVE

The density of the uniform representation of the Morgenstern[25] distribution is as follows

$$h_0(u,v) = 1 + \theta(2u-1)(2v-1), \quad |\theta| \leq 1.$$  

(4.1)

We have

$$E_{KS} = 1, \quad E_{KT} = 4/3.$$  

Recall that both rho and tau are locally most powerful and fully Pitman and Bahadur efficient under (4.1) (cf. Woodworth[31], Farlie[19] and Kremer[19]).
4.2. LIMITING PITMAN EFFICIENCIES UNDER THE WOODWORTH ALTERNATIVE

The density of the uniform representation of the family of distributions introduced in Woodworth[31] is the following

\[ h_{\phi,m}(u,v) = 1 + \phi(1-(m+1)u^m)(1-(m+1)v^m), \quad (4.2) \]

\[-1/m \leq \phi \leq 1/m^2, \quad m \geq 1.\]

For \( m = 1 \) this family coincides with (4.1). For every \( m \) and \( 0 \leq \phi \leq 1/m^2 \), formula (4.2) defines potted random variables. Moreover, we have

\[ E_{KS} = 1, \quad E_{KT} = (3/4)\{(m+1)/(m+2)\}^2 (m+1)^{2/m}. \]

Hence, if \( m \to \infty \) then \( E_{KT} \) and \( E_{KS} \) tend to 3/4.

4.3. LIMITING PITMAN EFFICIENCIES UNDER THE FARLIE ALTERNATIVE

The density of the uniform representation of the family of distributions proposed by Farlie[11] is as follows

\[ h_{\phi}(u,v) = 1 + \phi \text{sgn}\{(u-1/2)(v-1/2)\}, \quad |\phi| \leq 1. \]

In this case

\[ E_{KS} = 1, \quad E_{KT} = 3/4. \]
4.4. LIMITING PITMAN EFFICIENCIES UNDER THE MARSHALL-OLKIN ALTERNATIVE

The Marshall-Olkin[24] bivariate distribution is given by

\[ P(X > x, Y > y) = \exp\{ -\lambda_1 x - \lambda_2 y - \lambda_1 \lambda_2 \max(x, y) \} \]

\[ \lambda_1 \geq 0, \quad \lambda_2 \geq 0, \quad \lambda_1 \lambda_2 > 0. \]

The df of the uniform representation of (4.3) is of the form

\[ H_{\lambda_1, \lambda_2, \lambda_1 \lambda_2}(u, v) = u + v - 1 + (1 - u)^{1-a} (1 - v)^{1-b} \min\{(1-u)^a, (1-v)^b\}, \]

where \( a = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)}, \ b = \frac{\lambda_1 \lambda_2}{(\lambda_2 + \lambda_1)} \).

By Section 3.6 of Marshall and Olkin[24], the formula (4.3) defines p.q.d. random variables. They are independent iff \( \lambda_1 \lambda_2 = 0 \). So \( \lambda_1 \lambda_2 \) plays the role of \( \theta \) in this case. Besides, we have

\[ E_{KS} = 1, \ E_{KT} = \left( \frac{3}{16} \right) \left( (1+2s)^{1/s} + 2/(1+s)^2 \right), \ s = \frac{\lambda_1}{\lambda_2}. \]

If \( s = 1 \) then \( E_{KT} = 81/32 \). If \( s \to 0 \) then \( E_{KT} \to (3/16)e^2 \). Moreover, if \( s \to \infty \) then \( E_{KT} \to 3/4 \).

4.5. LIMITING PITMAN EFFICIENCIES UNDER THE KONIJN ALTERNATIVE
By Konijn[14], the positively $\theta$-dependent and uniformly distributed on $(0,1)$ random variables have df of the form

$$H_\theta(u,v) = (1-\theta)uv + \theta H_+ , \quad 0 \leq \theta \leq 1 ,$$

where $H_+$ is the upper Frechet's bound, i.e. $H_+(u,v) = \min(u,v)$. By Konijn[14], p.277,

$$B(H_\theta,K) = (\theta^2 + 2\theta)/2 , \quad B(H_\theta,S) = \theta .$$

Moreover,

$$B(H_\theta,T) = \{(12)^{1/2}\theta/2\} \sup_{0<p<1} p(1-p) = (12)^{1/2}\theta/8 .$$

Hence

$$E_{KS} = 1 , \quad E_{KT} = 4/3 .$$

REFERENCES


