isid/ms/2004/15
June 7, 2004
http://www.isid.ac.in/~statmath/eprints

Wilcoxon-signed rank test for associated sequences

Isha Dewan
B.L.S. Prakasa Rao

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

WILCOXON-SIGNED RANK TEST FOR ASSOCIATED SEQUENCES

Isha Dewan and B.L.S. Prakasa Rao
e-mail: isha@isid.ac.in , blsp@isid.ac.in
Indian Statistical Institute , 7, S.J.S. Sansanwal Marg
New Delhi 110016, INDIA

Abstract

Let $\{X_1, \ldots, X_n\}$ be stationary associated random variables with one dimensional marginal distribution function F . We study the properties of the classical sign statistic and the Wilcoxon-signed rank statistic for testing for shift in location in the above set up. In the process we extend the Newman's inequality to functions of bounded variation which are mixtures of absolutely continuous component and discrete component only.

Key words : Sign statistic, Wilcoxon-signed rank statistic, Associated random variables , Hoeffding identity.

1 Introduction

A finite family $\{X_1, ..., X_n\}$ of random variables is said to be *associated* if

$$\mathrm{Cov}(h_1(X_1, ..., X_n), h_2(X_1, ..., X_n)) \geq 0$$

for any two coordinatewise nondecreasing functions h_1, h_2 on R^n such that the covariance exists. An infinite family of random variables is said to be *associated* if every finite subfamily is associated (cf. Esary, Proschan and Walkup (1967)).

Let $\{X_n, n \geq 1\}$ be a stationary sequence of associated random variables. Let $F(x, \theta) = F(x - \theta),\ F \in \Omega_s$ where $\Omega_s = \{F : F(x) = 1 - F(-x)\}$ be the distribution function of X_1 and suppose that the distribution function F is absolutely continuous with bounded density function f .

Suppose the finite sequence of stationary associated random variables $\{X_i, 1 \leq i \leq n\}$ is observed. We wish to test the null hypothesis

$$H_0 : \theta = 0 \tag{1.1}$$

against the alternative hypothesis

$$H_1 : \theta > 0. \tag{1.2}$$

The most commonly used tests for this problem are the sign test and the Wilcoxon-signed rank test when the observations are independent and identically distributed. We now study the properties of these tests when the observations are on a stationary associated sequence of random variables.

Probabilistic aspects of associated random variables have been extensively studied (see, for example, Prakasa Rao and Dewan (2001) and Roussas(1999)). Dewan and Prakasa Rao (2003) studied the Wilcoxon Mann-Whitney statistic for stationary associated sequences.

In Section 2, we obtain a stochastic inequality for the covariance of functions $\alpha(X)$ and $\beta(Y)$ of a bivariate random vector (X, Y). This inequality follows from the results in Cuadras (2002). Of particular interest to us is the case when the random variables X and Y are associated. We discuss an application of this result to obtain tests for location for associated sequences in Section 3.

2 Stochastic Inequality

Let (X, Y) be a bivariate random vector and suppose that $E(X^2) < \infty$ and $E(Y^2) < \infty$. Further let

$$H(x, y) = P(X \leq x, Y \leq y) - P(X \leq x)P(Y \leq y). \tag{2.1}$$

Hoeffding (1940) proved that

$$\text{Cov}(X, Y) = \int_{R^2} H(x, y) \ \ dxdy, \tag{2.2}$$

hereafter called the *Hoeffding identity.* Lehmann (1966) gave a simple proof of this identity. Multivariate versions of this identity were studied by Jogdeo (1968) and Block and Fang (1988) using the concept of a cumulant of a random vector $\mathbf{X} = (X_1, \ldots, X_k)$. Yu (1993) obtained a generalisation of the covariance identity (2.2) to absolutely continuous functions of the components of the random vector $\mathbf{X}$ extending the earlier work of Newman (1984). Quesada-Molina (1992) generalised the Hoeffding identity to quasi-monotone functions $K(.,.)$ in the sense that

$$K(x, y) - K(x', y) - K(x, y') + K(x', y') \geq 0 \tag{2.3}$$

whenever $x \leq x'$ and $y \leq y'$. It was proved that

$$E[K(X, Y) - K(X^*, Y^*)] = \int_{R^2} H(x, y)K(dx, dy) \tag{2.4}$$

where X^* and $Y^*$are independent random variables independent of the random vector (X, Y) but with X^* and Y^* having the same marginal distributions as those of X and Y respectively. The results in Yu (1993) and Quesada-Molina (1992) were generalised to the multidimensional case in Prakasa Rao (1998). In a recent paper, Cuadras (2002) proved that if $\alpha(x)$ and $\beta(y)$ are functions of bounded variation on the support of the probabilty distribution of the random vector (X, Y) with $E|\alpha(X)\beta(Y)|, E|\alpha(X)|$ and $E|\beta(Y)|$ finite, then

$$\text{Cov}(\alpha(X), \beta(Y)) = \int_{R^2} H(x, y) \ \ \alpha(dx)\beta(dy). \tag{2.5}$$

It is clear that this result also follows as a special case of (2.4). It is possible to restate a multidimensional version of the result in (2.5) as a special case of the results in Prakasa Rao (1998).

Suppose that $\alpha(x)$ and $\beta(y)$ are functions of bounded variation which are mixtures of absolutely continuous component and discrete component only. Let $\alpha^{(c)}(x)$ and $\alpha^{(d)}(x)$ denote the absolutely continuous component and the discrete components of $\alpha(x)$ respectively. Let $x_i, i \geq 1$ be the jumps of $\alpha(x)$ with jump sizes $\alpha(x_i+0)-\alpha(x_i-0)=p_i \neq 0$. Similarly let $y_j, j \geq 1$ be the jumps of $\beta(y)$ with jump sizes $\beta(y_j+0)-\beta(y_j-0)=q_j \neq 0$. Further more let $\alpha^{(c)\prime}(x)$ denote the derivative of $\alpha^{(c)}(x)$ whenever it exists. Observe that the derivative of $\alpha^{(c)}(x)$ exists almost everywhere. Suppose that

$$\sup_x |\alpha^{(c)\prime}(x)| < \infty, \quad \sup_i |p_i| < \infty \tag{2.6}$$

and

$$\sup_y |\beta^{(c)\prime}(y)| < \infty, \quad \sup_j |q_j| < \infty. \tag{2.7}$$

Then

$$\begin{aligned}
\mathrm{Cov}(\alpha(X),\beta(Y)) &= \int_{R^2} H(x,y)\ \alpha(dx)\beta(dy) \\
&= \int_{R^2} H(x,y)\ \alpha^{(c)}(dx)\beta^{(c)}(dy) \\
&+ \int_{R^2} H(x,y)\ \alpha^{(c)}(dx)\beta^{(d)}(dy) \\
&+ \int_{R^2} H(x,y)\ \alpha^{(d)}(dx)\beta^{(c)}(dy) \\
&+ \int_{R^2} H(x,y)\ \alpha^{(d)}(dx)\beta^{(d)}(dy) \qquad (2.8)
\end{aligned}$$

and hence

$$\begin{aligned}
|\mathrm{Cov}(\alpha(X),\beta(Y))| \leq\ & \sup_x |\alpha^{(c)\prime}(x)| \sup_y |\beta^{(c)\prime}(y)| \int_{R^2} |H(x,y)|\ dxdy \\
& + \sup_x |\alpha^{(c)\prime}(x)| \sup_j |q_j| \sum_{j=1}^{\infty} \int_{-\infty}^{\infty} |H(x,y_j)| dx \\
& + \sup_y |\beta^{(c)\prime}(y)| \sup_i |p_i| \sum_{i=1}^{\infty} \int_{-\infty}^{\infty} |H(x_i,y)| dy \\
& + \sup_i |p_i| \sup_j |q_j| \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |H(x_i,y_j)|. \qquad (2.9)
\end{aligned}$$

Remark 2.1 : Suppose the functions $\alpha(x)$ and $\beta(y)$ are purely discrete. Let $x_i,\ i \geq 1$ be the jumps of $\alpha(x)$ with jump sizes $\alpha(x_i+0)-\alpha(x_i-0)=p_i \neq 0$ and $\alpha(x)$ is a constant between different jumps. Similarly let $y_j,\ j \geq 1$ be the jumps of $\beta(y)$ with jump sizes $\beta(y_j+0)-\beta(y_j-0)=q_j \neq 0$ and $\beta(y)$ is a constant between different jumps. Then

$$\mathrm{Cov}(\alpha(X),\beta(Y)) = \int_{R^2} H(x,y)\ \alpha^{(d)}(dx)\beta^{(d)}(dy)$$

$$= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} H(x_i, y_j) p_i q_j. \tag{2.10}$$

For instance, suppose that $\alpha(x) = sgn(x)$ and $\beta(y) = sgn(y)$ where $sgn(x) = 1$ if $x > 0$, $sgn(x) = -1$ if $x < 0$ and $sgn(x) = 0$ if $x = 0$. Then, for any x_0 and y_0,

$$\text{Cov}(sgn(X - x_0), sgn(Y - y_0)) = 4 \ \ H(x_0, y_0) \tag{2.11}$$

since the jump at x_0 is of size 2 for the function $\alpha(x)$ and for the function $\beta(y)$ (cf. Cuadras (2002)).

Remark 2.2 : Suppose we choose $\alpha(x) = F(x)$ and $\beta(y) = G(y)$ where $F(.)$ and $G(.)$ are continuous marginal distribution functions of the components X and Y respectively of a bivariate random vector (X, Y). Following the result given in (2.5), we get that

$$\text{Cov}(F(X), G(Y)) = \int_{R^2} H(x, y) \ \ F(dx)G(dy). \tag{2.12}$$

It is easy to see that the Spearman's rank correlation coefficient ρ between X and Y is given by

$$\rho = Corr(F(X), G(Y)) = 12 \ \ \text{Cov}(F(X), G(Y)) = 12 \int_{R^2} H(x, y) \ \ F(dx)G(dy). \tag{2.13}$$

Note that the random variables $F(X)$ and $G(Y)$ have the standard uniform distribution.

Remark 2.3: Suppose that X and Y are associated. Then $H(x, y) \geq 0$. Newman (1980) showed that if X and Y have finite variances, then, for any two differentiable functions h and g,

$$|\text{Cov}(h(X), g(Y))| \leq \sup_x |h'(x)| \ \ \sup_y |g'(y)| \ \ \text{Cov}(X, Y) \tag{2.14}$$

where h' and g' denote the derivatives of h and g, respectively. Inequality (2.4) extends this results to include functions of bounded variations which are mixtures of absolutely continuous component and discrete component only. Bulinski (1996) extended this result to continuous functions which are possibly not differentiable at atmost finite number of points. In the next section we discuss an application of the inequality to associated sequences.

3 Tests for Location for Associated Sequences

Here we study the sign test and the Wilcoxon signed rank test, based on the stationary associated random variables $\{X_i, 1 \leq i \leq n\}$, for testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$. Let C denote a generic positive constant in the sequel. Assume that

$$\sup_x f(x) < C. \tag{3.1}$$

Further assume that

$$\sum_{j=2}^{\infty} \text{Cov}^{\frac{1}{3}}(X_1, X_j) < \infty. \tag{3.2}$$

This would imply that $\mathrm{Cov}(X_1, X_n) \to 0$ as $n \to \infty$. In particular it follows that $\sup_n |\mathrm{Cov}(X_1, X_n)| < \infty$. Observing that $\mathrm{Cov}(X_1, X_n) > 0$ by associativity of $X_1, \ldots, X_n$, we obtain that

$$\begin{aligned} 0 &\leq \mathrm{Cov}(X_1, X_j) \\ &= [\mathrm{Cov}(X_1, X_j)]^{\frac{2}{3}}[\mathrm{Cov}(X_1, X_j)]^{\frac{1}{3}} \\ &\leq [\sup_n \mathrm{Cov}(X_1, X_n)]^{\frac{2}{3}}[\mathrm{Cov}(X_1, X_j)]^{\frac{1}{3}}. \end{aligned}$$

Hence

$$\sum_{j=2}^{\infty} \mathrm{Cov}(X_1, X_j) \leq [\sup_n \mathrm{Cov}(X_1, X_n)]^{\frac{2}{3}} \sum_{j=2}^{\infty} [\mathrm{Cov}(X_1, X_j)]^{\frac{1}{3}} < \infty. \tag{3.3}$$

3.1 The Sign Test

For testing the hypothesis $H_0 : \theta = 0$ against $H_1 : \theta > 0$, the sign test is based on the statistic

$$U_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i), \tag{3.4}$$

where

$$\phi(x) = I(x > 0), \tag{3.5}$$

and $I(A)$ denotes the indicator function of the set A. Observe that

$$\begin{aligned} E(\phi(X_1)) &= 1 - F(0) = p \ \ \text{(say)}, \\ \mathrm{Var}(\phi(X_1)) &= p - p^2, \end{aligned}$$

and

$$\mathrm{Cov}(\phi(X_1), \phi(X_j)) = P[X_1 > 0, X_j > 0] - p^2. \tag{3.6}$$

Since the density function $f(.)$ of the random variable X_1 is bounded, it follows,from Bagai and Prakasa Rao (1991), that

$$\sup_{x,y} |P[X_1 > x, Y_1 > y] - P[X_1 > x]P[Y_1 > y]| \leq C \ \ \mathrm{Cov}^{1/3}(X, Y). \tag{3.7}$$

From (2.10), (3.3) and (3.7) it follows that

$$\sigma^2 = \mathrm{Var}\ \phi(X_1) + 2 \sum_{j=2}^{\infty} \mathrm{Cov}(\phi(X_1), \phi(X_j)) < \infty. \tag{3.8}$$

Since $\phi(x)$ is an increasing function of x, we have $\phi(X_1), \phi(X_2), \ldots, \phi(X_n)$ are stationary associated random variables. The following theorem is an immediate consequence of the central limit theorem for associated random variables (Newman (1980)).

Theorem 3.1 Let $X_n, n \geq 1$ be a sequence of stationary associated random variables with bounded density function. Then

$$\frac{n^{-1/2} \sum_{j=1}^{n} [\phi(X_j) - E(\phi(X_j))]}{\sigma} \xrightarrow{\mathcal{L}} N(0,1) \ \ \text{as} \ \ n \to \infty. \tag{3.9}$$

The test procedure consists in rejecting the null hypothesis H_0 for large values of the statistic $U_n^{(1)}$, which is the proportion of positive observations.

3.2 Wilcoxon-signed Rank Test

Let $R_1, R_2, \ldots, R_n$ be the ranks of $X_1, X_2, \ldots, X_n$. The Wilcoxon-signed rank statistic is defined by

$$T = \sum_{j=1}^{n} R_j \phi(X_j). \tag{3.10}$$

We can write T as a linear combination of two U-statistics (Hettmansperger (1984))

$$T = nU_n^{(1)} + \binom{n}{2} U_n^{(2)}, \tag{3.11}$$

where

$$\binom{n}{2} U_n^{(2)} = \sum_{1 \le i < j \le n} \psi(X_1, X_j), \tag{3.12}$$

and

$$\psi(x, y) = I(x + y > 0). \tag{3.13}$$

Since the random variables $\{X_n, n \ge 1\}$ form a stationary sequence, it follows that

$$\begin{aligned} E(U_n^{(2)}) &= \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} p_{ij} \\ &= \frac{1}{\binom{n}{2}} \sum_{j=2}^{n} (n - j + 1) p_{1j} \end{aligned} \tag{3.14}$$

where $p_{ij} = P[X_i + X_j > 0]$. Let

$$\begin{aligned} \theta &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x, y) dF(x) dF(y) \ , \\ &= 1 - \int_{-\infty}^{\infty} F(-x) dF(x), & (3.15) \\ \psi_1(x_1) &= E(\psi(x_1, X_2)) \\ &= \int_{-\infty}^{\infty} \psi_(x_1, x_2) dF(x_2) \\ &= 1 - F(-x_1). & (3.16) \\ h^{(1)}(x_1) &= \psi_1(x_1) - \theta \ , & (3.17) \end{aligned}$$

and

$$\begin{aligned} h^{(2)}(x_1, x_2) &= \psi(x_1, x_2) - \psi_1(x_1) - \psi_1(x_2) + \theta \\ &= \psi(x_1, x_2) + F(-x_1) + F(-x_2) - 2 + \theta. \end{aligned} \tag{3.18}$$

Then the Hoeffding-decomposition (H-decomposition) for $U_n^{(2)}$ is given by (see, Lee (1990))

$$U_n^{(2)} = \theta + 2H_n^{(1)} + H_n^{(2)} \tag{3.19}$$

where $H_n^{(j)}$ is a U-statistic of degree j based on the kernel $h^{(j)}$, $j = 1, 2$, that is,

$$H_n^{(j)} = \frac{1}{\binom{n}{j}} \sum h^{(j)}(X_{i_1}, \ldots, X_{i_j}) \tag{3.20}$$

where summation is taken over all subsets $1 \leq i_1 < \ldots < i_j \leq n$ of $\{1, \ldots, n\}$.

In view of the H - decomposition , we have

$$\mathrm{Var}(U_n^{(2)}) = 4\ \mathrm{Var}(H_n^{(1)}) + \mathrm{Var}(H_n^{(2)}) + 4\ \mathrm{Cov}(H_n^{(1)}, H_n^{(2)}). \tag{3.21}$$

From Dewan and Prakasa Rao (2001), we get that

$$\mathrm{Var}(H_n^{(1)}) = \frac{1}{n}(\sigma_1^2 + 2\sum_{j=2}^{\infty} \sigma_{1j}^2) + o(\frac{1}{n}), \tag{3.22}$$

where

$$\begin{aligned} \sigma_1^2 &= \mathrm{Var}(F(-X_1)) \\ \sigma_{1j}^2 &= \mathrm{Cov}(F(-X_1), F(-X_{1+j})) \end{aligned} \tag{3.23}$$

Using the Newman's inequality and the inequality (3.3) , we get

$$\sum_{j=2}^{\infty} \sigma_{1j}^2 = \sum_{j=2}^{\infty} \mathrm{Cov}(F(-X_1), F(-X_{1+j})) < \infty \tag{3.24}$$

Furthermore

$$\mathrm{Var}(H_n^{(2)}) = \binom{n}{2}^{-2} \sum_{1 \leq i < j \leq n} \sum_{1 \leq k < \ell \leq n} \mathrm{Cov}\{h^{(2)}(X_i, X_j), h^{(2)}(X_k, X_\ell)\} \tag{3.25}$$

where

$$\begin{aligned} \mathrm{Cov}\{h^{(2)}(X_i, X_j), h^{(2)}(X_k, X_\ell)\} &= \mathrm{Cov}(\psi(X_i, X_j), \psi(X_k, X_\ell)) \\ &+ \mathrm{Cov}(\psi(X_i, X_j), F(-X_k)) + \mathrm{Cov}(\psi(X_i, X_j), F(-X_\ell)) \\ &+ \mathrm{Cov}(\psi(X_k, X_\ell), F(-X_i)) + \mathrm{Cov}(\psi(X_k, X_\ell), F(-X_j)) \\ &+ \mathrm{Cov}(F(-X_i), F(-X_k)) + \mathrm{Cov}(F(-X_i), F(-X_\ell)) \\ &+ \mathrm{Cov}(F(-X_j), F(-X_k)) + \mathrm{Cov}(F(-X_j), F(-X_\ell)). \end{aligned} \tag{3.26}$$

Using the Newman's (1980) inequality, we get that

$$|\mathrm{Cov}(F(-X_i), F(-X_k))| \leq \sup_x (f(x))^2 \mathrm{Cov}(X_i, X_k). \tag{3.27}$$

Since the density function is bounded, it follows, from Bagai and Prakasa Rao (1991), that

$$\begin{aligned} &|\mathrm{Cov}(\psi(X_i, X_j), \psi(X_k, X_\ell))| \\ &= |P[X_i + X_j > 0, X_\ell + X_k > 0] - P[X_i + X_j > 0]P[X_\ell + X_k > 0]| \\ &\leq C[\mathrm{Cov}(X_i + X_j, X_k + X_\ell)]^{1/3} \\ &= C[\mathrm{Cov}(X_i, X_k) + \mathrm{Cov}(X_j, X_k) + \mathrm{Cov}(X_i, X_\ell) + \mathrm{Cov}(X_j, X_\ell)]^{1/3} \end{aligned} \tag{3.28}$$

Let $Z = X_i + X_j$. Note that the function $\psi(x_i, x_j) = I(x_i + x_j > 0) = I(z > 0)$ has a jump of size 1 at $z = 0$. Then, from (2.5), it follows that

$$
\begin{aligned}
&|\mathrm{Cov}(\psi(X_i, X_j), F(-X_k))| \\
&= |\int_{-\infty}^{\infty} P[X_i + X_j \le 0, X_k \le x] - P[X_i + X_j \le 0] P[X_k \le x]) dF(x)| \\
&\le \int_{-\infty}^{\infty} |P[X_i + X_j \le 0, X_k \le x] - P[X_i + X_j \le 0] P[X_k \le x]| dF(x) \\
&\le C \int_{-\infty}^{\infty} [\mathrm{Cov}(X_i + X_j, X_k)]^{1/3} dF(x) \\
&= C[\mathrm{Cov}(X_i + X_j, X_k)]^{1/3} \\
&= C[\mathrm{Cov}(X_i, X_k) + \mathrm{Cov}(X_j, X_k)]^{1/3}. \qquad (3.29)
\end{aligned}
$$

Using (3.27), (3.28) and (3.29) in (3.26), we get

$$
\begin{aligned}
&|\mathrm{Cov}\{h^{(2)}(X_i, X_j), h^{(2)}(X_k, X_\ell)\}| \\
&\le C[\mathrm{Cov}(X_i, X_k) + \mathrm{Cov}(X_j, X_k) + \mathrm{Cov}(X_i, X_\ell) + \mathrm{Cov}(X_j, X_\ell)]^{1/3} \\
&+ [\mathrm{Cov}(X_i, X_k) + \mathrm{Cov}(X_j, X_k)]^{1/3} + [\mathrm{Cov}(X_i, X_\ell) + \mathrm{Cov}(X_j, X_\ell)]^{1/3} \\
&+ [\mathrm{Cov}(X_k, X_i) + \mathrm{Cov}(X_\ell, X_i)]^{1/3} + [\mathrm{Cov}(X_k, X_j) + \mathrm{Cov}(X_\ell, X_j)]^{1/3} \\
&+ \mathrm{Cov}(X_i, X_k) + \mathrm{Cov}(X_j, X_k) + \mathrm{Cov}(X_i, X_\ell) + \mathrm{Cov}(X_j, X_\ell) \\
&\le C[\mathrm{Cov}(X_i, X_k) + \mathrm{Cov}(X_j, X_k) + \mathrm{Cov}(X_i, X_\ell) + \mathrm{Cov}(X_j, X_\ell)] \\
&+ C[\mathrm{Cov}(X_i, X_k)^{1/3} + \mathrm{Cov}(X_j, X_k)^{1/3} + \mathrm{Cov}(X_i, X_\ell)^{1/3} + \mathrm{Cov}(X_j, X_\ell)^{1/3}] \\
&= C[(\mathrm{Cov}(X_i, X_k) + \mathrm{Cov}(X_i, X_k)^{1/3}) + (\mathrm{Cov}(X_j, X_k) + \mathrm{Cov}(X_j, X_k)^{1/3}) \\
&+ (\mathrm{Cov}(X_i, X_\ell) + \mathrm{Cov}(X_i, X_\ell)^{1/3}) + (\mathrm{Cov}(X_j, X_\ell) + \mathrm{Cov}(X_j, X_\ell)^{1/3}) \\
&= r(|i-k|) + r(|j-k|) + r(|i-\ell|) + r(|j-\ell|) \text{ (say)}. \qquad (3.30)
\end{aligned}
$$

From (3.2) and (3.3) it follows that

$$\sum_{k=1}^{\infty} r(k) < \infty. \qquad (3.31)$$

Hence, following Serfling (1968), we have , as $n \to \infty$

$$\mathrm{Var}(H_n^{(2)}) = o(\frac{1}{n}). \qquad (3.32)$$

Using the Cauchy-Schwartz inequality, it follows that

$$\mathrm{Cov}(H_n^{(1)}, H_n^{(2)}) = o(\frac{1}{n}). \qquad (3.33)$$

From (3.21), (3.22), (3.32) and (3.33), we get that

$$\mathrm{Var}(U_n^{(2)}) = 4[\sigma_1^2 + 2\sum_{j=1}^{\infty} \sigma_{1j}^2] + o(\frac{1}{n}). \qquad (3.34)$$

Then using the same technique as in Theorem 3.2 in Dewan and Prakasa Rao (2001) for obtaining the limiting distributiion of U-statistics , we get the following theorem.

Theorem 3.2: Let $\{X_n, n \geq 1\}$ be an associated sequence . Suppose (3.2) holds. Then

$$\frac{n^{1/2}(U_n^{(2)} - \theta)}{2\sigma_U} \xrightarrow{\mathcal{L}} N(0,1) \text{ as } n \to \infty \tag{3.35}$$

where $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty} \sigma_{1j}^2$.

Finally, define

$$T^* = \frac{T - \gamma}{\binom{n}{2}}, \tag{3.36}$$

where

$$\gamma = nP[X > 0] + \binom{n}{2}\theta. \tag{3.37}$$

The following theorem gives the limiting distribution of the Wilcoxon signed rank statistic.

Theorem 3.3: Let $\{X_n, n \geq 1\}$ be an associated sequence with a bounded density function. Suppose (3.2) holds. Then as

$$\frac{n^{1/2}T^*}{2\sigma_U} \xrightarrow{\mathcal{L}} N(0,1) \text{ as } n \to \infty. \tag{3.38}$$

Proof : Note that

$$E[U_n^{(1)}] = P[X > 0], \tag{3.39}$$

and from (3.8)

$$\frac{1}{n}\text{Var}[U_n^{(1)}] \to 0 \text{ as } n \to \infty. \tag{3.40}$$

The result now follows using Theorem 3.2 and Slutsky's theorem.

Note that the Wilcoxon signed rank statistic T is the sum of ranks of positive observations. The test procedure consists in rejecting the null hypothesis H_0 for large values of the statistic T. The quantity σ_U^2 depends on the unknown distribution F even under the null hypothesis. It can be estimated using the estimators given by Peligrad and Suresh (1995) and the result of Roussas (1993) (cf. Dewan and Prakasa Rao (2003)) for estimating the variance of Wilcoxon Mann-Whitney statistic for associated sequences) . A consistent estimator of σ_U^2 is given by

$$J_n^2 = \frac{\pi}{2}\hat{B}_n^2, \tag{3.41}$$

where, for $\ell = \ell_n$,

$$\hat{B}_n^2 = \frac{1}{n-\ell}\sum_{j=0}^{n-\ell} \frac{|\hat{S}_j(\ell) - \ell\hat{\bar{\psi}}_n|}{\sqrt{\ell}}, \tag{3.42}$$

and $\hat{S}_j(k) = \sum_{i=j+1}^{j+k} \hat{\psi}_1(X_i)$, $\hat{\bar{\psi}}_n = \frac{1}{n}\sum_{i=1}^{n} \hat{\psi}_1(X_i)$, $\hat{\psi}_1(x) = 1 - F_n(-x)$, where F_n is the empirical distribution function corresponding to F based on associated random variables $X_1, X_2, \ldots, X_n$. Note that under the null hypothesis X and $-X$ are identically distributed.

References

Block, H.W. and Fang Z. (1988), A multivariate extension of Hoeffding's Lemma, *Ann. Probab.*, **16**, 1803-1820.

Bulinski, A.V. (1996). On the convergence rates in the CLT for positively and negatively dependent random fields, in: I.A. Ibragimov and A.Yu. Zaitsev, eds., *Probability Theory and Mathematical Statistics*, Gordon and Breach, U.K.

Cuadras, C.M. (2002). On the covariance between functions, *J. Mult. Anal.*, **81**, 19-27.

Dewan, I. and Prakasa Rao, B.L.S., (2001). Asymptotic normality of U-statistics of associated random variables, *J. Statist. Plan. Inference*, **97**, 201-225.

Dewan, I. and Prakasa Rao, B.L.S., (2003). Mann-Whitney test for associated sequences, *Ann. Inst. Statist. Math.*, **55**, 111-120.

Esary, J., Proschan, F. and Walkup, D. (1967). Association of random variables with applications, *Ann. Math. Statist.*, **38**, 1466-1474.

Hettmansperger, T.P. (1984). *Statistical inference based on ranks*, Wiley, New York.

Hoeffding, W. (1940). Masstabinvariante Korrelations-theorie, *Schr. Math. Inst., University Berlin*, **5**, 181-233.

Jogdeo, K. (1968). Characterization of independence in certain families of bivariate and multivariate distributions., *Ann. Math. Statist.* , **39**, 433-441.

Lee, A.J. (1990). *U-Statistics*, Marcel Dekker, New York.

Lehmann, E.L. (1966). Some concepts of dependence, *Ann. Math. Statist.*, **37**, 1137-1153.

Newman, C.M. (1980). Normal fluctuations and the FKG inequalities, *Comm. Math. Phys.*, **74**, 119-128.

Newman, C.M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In : *Inequalities in Statistics and Probability*, Tong, Y.L. (ed.), 127-140, IMS, Hayward.

Peligard, M. and Suresh, R. (1995). Estimation of variance of partial sums of an associated sequence of random variables, *Stoch. Process. Appl.* , **56**, 307-319.

Prakasa Rao, B.L.S. (1998). Hoeffding identity, multivariance and multicorrelation, *Statistics*, **32**, 13-29.

Prakasa Rao. B.L.S. and Dewan, I. (2001). Associated sequences and related inference problems, *Handbook of Statistics, 19 , Stochastic Processes: Theory and Methods*, (eds. C.R. Rao and D.N. Shanbhag), 693-728, North Holland, Amsterdam.

Quesada- Molina , J.J. (1992). A generalization of an identity of Hoeffding and some applications, *J. Ital, Stat. Sc.*, **1**, 405-411.

Roussas, G.G. (1993) . Curve estimation in random field of associated processes, *J. Nonparametric Statist.*, **2**, 215-224.

Roussas, G.G. (1999) . Positive and negative dependence with some statistical applications, *Asymptotics, nonparametrics and time series* (ed. S. Ghosh), 757-788, Marcel Dekker, New York.

Serfling , R.J. (1968). The Wilcoxon two-sample statistic on strongly mixing processes, *Ann. Math. Statist.*, **39**, 1202-1209.

Yu, H. (1993). A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences, *Probab. Th. Rel. Fields*, **95**, 357-370.