

# A Class of Distribution-free Tests for Independence against Positive Quadrant Dependence

Parameshwar V Pandit  
Department of Statistics  
Bangalore University  
Bangalore -560056, India  
panditpv@rediffmail.com

Savitha Kumari  
Department of Statistics  
SDM College, Ujire, India  
savi\_rrao@yahoo.co.in

## Abstract

A class of distribution-free tests based on convex combination of two U-statistics is considered for testing independence against positive quadrant dependence. The class of tests proposed by Kochhar and Gupta (1987) and Kendall's test are members of the proposed class. The performance of the proposed class is evaluated in terms of Pitman asymptotic relative efficiency for Block- Basu (1974) model and Woodworth family of distributions. It has been observed that some members of the class perform better than the existing tests in the literature. Unbiasedness and consistency of the proposed class of tests have been established.

**Keywords & Phrases:** Distribution-free, Positive quadrant dependence, Convex combination of test statistics, U-statistic.

## 1. Introduction

Dependence relations between random variables is one of the most widely studied subjects in Probability and statistics. Many studies in statistics are designed to explore the relationship between random variables  $X$  and  $Y$ , say and specially to determine whether  $X$  and  $Y$  are independent or dependent. For example, a doctor may be interested in studying the relationship between obesity and blood pressure. Specifically, he may be interested in testing whether obesity and blood pressure are independent against the alternative that they are positively associated. Here  $X$  could be the degree of overweight as measured by the ratio of actual body weight to ideal body weight as given in certain standard tables, and  $Y$  could be the systolic blood pressure. Similarly a meat vender may be interested in knowing whether colour and quality of canned tuna are independent or the consumer preference is towards light tuna. A psychologist may be interested in testing the hypothesis that the time until a it takes an infant to walk alone is independent of the infant's IQ at a later age versus the alternative that children who learn to walk early tend to have higher IQs. An engineer may be interested in the dependence relationship between system reliability and the environment. Specifically he may be interested in the positive association between system reliability and a certain characteristic of the environment.

Let  $(X, Y)$  be absolutely continuous random variable with joint distribution function  $F(x, y)$  and survival function  $\bar{F}(x, y) = P[X > x, Y > y]$ . Let  $F$  and  $G$  denote the marginal distribution functions of  $X$  and  $Y$  respectively with corresponding survival function  $\bar{F} = 1 - F$  and  $\bar{G} = 1 - G$

Definition:  $(X, Y)$  is positive quadrant dependent (PQD) if  $F(x, y) \geq F(x)G(y)$ , for all  $(x, y)$ . or  $\bar{F}(x, y) \geq P[X > x, Y > y]$ , for all  $(x, y)$

The dependence is strict if the inequality holds at least for one pair  $(x, y)$ . The concept of PQD is due to Lehmann (1996)

A measure of general positive association between two random variables  $X$  and  $Y$  is defined in terms of covariance between every pair of non decreasing real function  $f$  and  $g$  as  $Cov[f(X), g(Y)] \geq 0$

However if a pair  $(X, Y)$  is PQD then  $Cov(X, Y) \geq 0$  equality holding if  $X, Y$  are independent. Further if  $f, g$  is a pair of real and non decreasing functions, then  $(X, Y)$  is PQD implies that  $[f(X), g(Y)]$  is PQD which in turn implies that  $Cov[f(X), g(Y)] \geq 0$ . Consequently general positive association and PQD are equivalent.

The concept of PQD has been used to construct conservative confidence intervals for the components of the mean vector in bivariate normal distribution. Many applications of this concept may be founding the study of contaminated independence models, slippage problems, tests of symmetry etc.

In this paper, we consider the problem of testing the null hypothesis of independence  $H_0 : F(x, y) = F(x)G(y)$  for all  $(x, y)$  against the alternative of PQD  $H_1 : F(x, y) \geq F(x)G(y)$  for all  $(x, y)$  with strict inequality holding on a set of non-zero probability.

It should be noted here that since  $F$  and  $G$  are unknown,  $H_0$  and  $H_1$  are both composite. Moreover the alternative  $H_1$  are ordered alternatives and this gives rise to general difficulties associated with ordered restricted inference.

Kochar and Gupta (1987, 1990) proposed some competitors of Kendall's sample tau coefficient for testing independence against PQD. Shriever (1987) contains a list of large number of tests available in the literature for the problem of independence. Shetty and Pandit (1996, 1998, 2003) proposed distribution-free tests for this problem based on the ordering of observations in sub-samples. The statistic proposed by Kochar and Gupta (1990) is a member of the test proposed by Shetty and Pandit (2003). In this paper, we propose a class of distribution-free tests for this problem.

In section 2 we propose a new class of distribution – free tests based on U- statistics for testing  $H_0$  against  $H_1$ . The distribution of the test statistics are considered in Section 3. Section 4 is devoted to asymptotic relative efficiency properties.

## 2. The Proposed Test

On the basis of random sample  $(x_1, y_1), \dots, (x_n, y_n)$  from the distribution  $H$ , we wish to test  $H_0$  against  $H_1$ . Let  $k \geq 2$  be a fixed positive integer and consider the following kernels.

$$\Phi_{1k}[(x_1, y_1), \dots, (x_k, y_k)] = \begin{cases} 1 & \text{if } \text{Max}(x_1, \dots, x_k) \text{ and } \text{Max}(y_1, \dots, y_k) \text{ belongs to the same pair} \\ 0 & \text{Otherwise} \end{cases}$$

$$\Phi_{2k}[(x_1, y_1), \dots, (x_k, y_k)] = \begin{cases} 1 & \text{if } \text{Min}(x_1, \dots, x_k) \text{ and } \text{Min}(y_1, \dots, y_k) \text{ belongs to the same pair} \\ 0 & \text{Otherwise} \end{cases}$$

Let

Let  $\Phi_k = \Phi_{1k} + \Phi_{2k}$  and  $U_k$  be the  $U$  statistic corresponding to the kernel  $\Phi_k$  defined by

$$U_k = \binom{n}{k}^{-1} \sum \Phi_k[(X_{i_1}, Y_{i_1}), \dots, (X_{i_k}, Y_{i_k})],$$

where the summation is over all combinations of  $k$  integers  $(i_1, i_2, \dots, i_k)$  chosen out of  $n$  integers  $(1, 2, \dots, n)$ .

Large values of  $U_k$  are significant for testing  $H_0$  against  $H_1$ . Obviously  $U_2$  is the celebrated Kendall's tau statistic. Kochar and Gupta (1987) have proposed tests based on  $U$  statistics associated with kernels  $\Phi_{1k}$  for the above problem in the case of skewed alternatives. It is expected that the newly proposed tests will be quite efficient for the symmetric case.

The statistic  $U_k$  can be easily computed as follows. Assume without loss of generality that  $X_1 \leq \dots \leq X_n$  and denote the corresponding  $Y$ s by  $Y_{[1]}, \dots, Y_{[n]}$ .

Let  $L_{(j)} = \text{rank of } Y_{(j)} \text{ among } Y_{[1]}, \dots, Y_{[j]}$ . and  $M_{(j)} = \text{number of } Y_{[i]} \text{ s greater than or equal to } Y_{[j]}, Y_{[j+1]} \dots, Y_{[n]}$ . Then  $U_k$  can be written as

$$U_k = \frac{\alpha \sum_{j=1}^n \binom{L_{(j)} - 1}{k - 1} + (1 - \alpha) \sum_{j=1}^n \binom{M_{(j)} - 1}{k - 1}}{\binom{n}{k}}$$

Woodworth (1965) calls  $L_{(j)}$  the third quadrant layer rank of  $(X_j, Y_{[j]})$ .

## 3. Distribution of $U_k$

The expectation of  $U_k$  is given by

$$\begin{aligned} E[U_k] &= k[\alpha P(X_i \leq X_j, Y_i \leq Y_j \text{ for } i = 1, 2, \dots, k, i \neq j) + \{1 - \alpha\} P(X_j \leq X_i, Y_j \leq Y_i \text{ for } i = 1, 2, \dots, k, i \neq j)] \\ &= k \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} [\alpha F^{k-1}(x, y) + (1 - \alpha) \bar{F}^{k-1}(x, y)] dF(x, y) \\ &\equiv \tau_{1k} + \tau_{2k} \end{aligned}$$

Under  $H_0$ ,  $E[U_k] = 1/k$ . But under the alternative  $H_1$

$$\begin{aligned}
 E[U_k] &= k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\alpha F^{k-1}(x, y) + (1 - \alpha) \bar{F}^{k-1}(x, y)] dF(x, y) \\
 &> k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\alpha F^{k-1}(x, y) G^{k-1}(x, y) + (1 - \alpha) \bar{F}^{k-1}(x, y) \bar{G}^{k-1}(x, y)] dF(x, y) \\
 &= k [\alpha P\{\max(X_{1, \dots}, X_{k-1}) \leq X_k, \max(Y_{k+1, \dots}, Y_{2k-1}) \leq Y_k\} + (1 - \alpha) P\{\min(X_{1, \dots}, X_{k-1}) \geq X_k, \min(Y_{k+1, \dots}, Y_{2k-1}) \geq Y_k\}] \\
 &= k [\alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\bar{F}(x, y) dF^{k-1}(x) dG^{k-1}(y) + (1 - \alpha) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x, y) d\bar{F}^{k-1}(x) d\bar{G}^{k-1}(y)] \\
 &> k [\alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\bar{F}(x) \bar{G}(y) dF^{k-1}(x) dG^{k-1}(y) + (1 - \alpha) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x) G(y) d\bar{F}^{k-1}(x) d\bar{G}^{k-1}(y)] = 1/k
 \end{aligned}$$

### Asymptotic distribution of $U_k$

The asymptotic distribution of  $\sqrt{n}[U_k - E(U_k)]$  as  $n \rightarrow \infty$  is normal with mean zero and variance  $k^2 \xi_{1k}$ , where

$$\xi_{1k} = E[\Psi_{1k}^2(X_1, Y_1)] - E^2(U_k)$$

$$\begin{aligned}
 \text{Where } \Psi_{1k}(X_1, Y_1) &= E\{\Phi_k[(x_1, y_1), \dots, (x_k, y_k)]\} \\
 &= E\{\alpha \Phi_{1k}[(x_1, y_1), \dots, (x_k, y_k)] + (1 - \alpha) \{\Phi_{2k}[(x_1, y_1), \dots, (x_k, y_k)]\}\} \\
 &= \alpha E[\Phi_{1k}] + (1 - \alpha) E[\Phi_{2k}] \\
 &= \alpha \{F^{k-1}(x_1) G^{k-1}(x_1) + \left(\frac{1}{k} - 1\right) [1 - F^{k-1}(x_1)] [1 - G^{k-1}(x_1)]\} \\
 &+ (1 - \alpha) \{\bar{F}^{k-1}(x_1) \bar{G}^{k-1}(y_1) + \left(\frac{1}{k} - 1\right) [1 - \bar{F}^{k-1}(x_1)] [1 - \bar{G}^{k-1}(y_1)]\} \\
 &= \alpha I_1 + (1 - \alpha) I_2
 \end{aligned}$$

Under  $H_0$ ,  $E[U_k] = 1/k$  and

$$\sigma_k^2 = k^2 \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\alpha I_1 + \{1 - \alpha\} I_2)^2 dF(x_1) dG(y_1) - (1/k)^2 \right]$$

Under  $H_0$ ,  $\sigma_k^2 = 1/9$ ,  $[15 \alpha^2 - 15 \alpha + 32]/200$ ,  $[2128 \alpha^2 - 2128 \alpha + 2025]/11025$ ,  $[35535 \alpha^2 - 35535 \alpha + 25088]/127008$  for  $k = 2, 3, 4$  respectively.

### 4. Asymptotic Relative Efficiencies

For asymptotic relative efficiency comparisons we consider the Woodworth (1966) family of distributions. The uniform representation of the distribution is

$$f(x, y) = 1 + \theta [1 - (m + 1)x^m] [1 - (m + 1)y^m], 0 \leq \theta \leq 1/m^2, m \geq 1$$

It is evident that for every  $m$  and the above choice of  $\theta$ , above function defines PQD random variables. For  $m = 1$ , coincides with the well known Morgenstern (1956) distribution,  $f(x, y) = 1 + \theta(1 - 2x)(1 - 2y), 0 \leq x \leq 1, 0 \leq y \leq 1, \theta \geq 0$

For which Kendall's tau and Spearman's rho tests are known to be asymptotically optimal.

The asymptotic relative efficiencies (AREs) of the  $U_k$  tests, relative to Kendall's tau test for different values of  $k$  are presented in table 1.

**Table 1: Asymptotic Relative Efficiencies of  $U_k$  relative to Kentall's tau test**

$\alpha$	k	m=2	m=3	m=4	m=5
0	3	0.6670	0.5487	0.4734	0.4238
	4	0.4405	0.3191	0.2531	0.2124
	5	0.3836	0.2016	0.1513	0.1226
0.1	3	0.7397	0.6391	0.5753	0.5315
	4	0.5428	0.4349	0.3766	0.3414
	5	0.4043	0.3115	0.2679	0.2448
0.2	3	0.8121	0.7341	0.6848	0.6511
	4	0.6555	0.5724	0.5314	0.5095
	5	0.5243	0.4557	0.4318	0.4263
0.3	3	0.8825	0.8314	0.8004	0.7799
	4	0.7709	0.7268	0.7138	0.7142
	5	0.6574	0.6312	0.6434	0.6701
0.4	3	0.9465	0.9277	0.9189	0.9148
	4	0.8839	0.8886	0.9141	0.9462
	5	0.7968	0.8256	0.8907	0.9650
0.5	3	1.0044	1.0223	1.0366	1.0516
	4	0.9808	1.0449	1.1176	1.1888
	5	0.9068	1.0164	1.1483	1.2824
0.6	3	1.0529	1.1059	1.1498	1.1862
	4	1.0539	1.1823	1.3072	1.4222
	5	0.9895	1.1806	1.3835	1.5826
0.7	3	1.0906	1.1823	1.3170	1.3143
	4	1.0979	1.2904	1.4681	1.6279
	5	1.0291	1.2983	1.5694	1.8306
0.8	3	1.1173	1.2472	1.3494	1.4323
	4	1.1134	1.3644	1.5915	1.7933
	5	1.0332	1.3646	1.6936	2.0076
0.9	3	1.1333	1.2997	1.4314	1.5375
	4	1.1038	1.4051	1.6751	1.9141
	5	1.0056	1.3858	1.7591	2.1136
1	3	1.1378	1.3396	1.4994	1.6384
	4	1.0754	1.4173	1.7226	1.9924
	5	0.9596	1.3732	1.7777	2.1608

We also study the ARE's of the newly proposed tests for the absolutely continuous bivariate exponential distribution of Block & Basu (1974) with density function.

$$f(x, y) = \frac{1}{2}(\theta + 1)(\theta + 2) \exp[-\{\min(x, y) + (1 + \theta) \max(x, y)\}], x, y \geq 0$$

This distribution is PQD when  $\theta > 0$  and the variables are independent when  $\theta = 0$ . The values of the AREs of the  $U_k$  tests with respect to the Kendall's tau test are presented in table 2 below:

**Table 2: Asymptotic Relative Efficiencies of  $U_k$**

$\alpha$	k=3	k=4	
	ARE( $U_k, T_n$ )	ARE( $U_k, T_n$ )	ARE( $U_k, KG_4$ )
0	0.6944	0.4900	0.4161
0.1	0.7696	0.6024	0.5115
0.2	0.8432	0.7257	0.6163
0.3	0.9152	0.8534	0.7247
0.4	0.9808	0.9755	0.8284
0.5	1.0400	1.0806	0.9179
0.6	1.0896	1.1594	0.9845
0.7	1.1280	1.2064	1.0244
0.8	1.1536	1.2217	1.0375
0.9	1.1696	1.2099	1.0274
1	1.1744	1.1776	1.0000

**5. Some Remarks and Conclusions:**

1. The proposed class of tests is unbiased and consistent for testing against  $H_0$  against  $H_1$ .
2. The asymptotic relative efficiencies (AREs) of the members of the class is studied for two alternatives, namely, Woodwoth family and Block-Basu bivariate exponential distribution.
3. The AREs of the proposed test relative to Kendall's tau test when the alternative is Woodworth's family is evaluated for different values of m.
4. It is observed that for  $m=3, 4$  and  $5$ , the AREs of proposed test relative to Kendall's tau test is more than one for those values of  $\alpha \geq 0.5$ , indicating the better performance of members of new class of tests. However, for  $m=2$ , the performance is better for  $\alpha \geq 0.6$ . The sub sample sizes considered are 3, 4 and 5.
5. For the sub sample sizes 3 and 4, the performances of the members of the class of test is better than Kendall's tau test for  $\alpha \geq 0.5$  and is better than the test due to Kochar and Gupta(1990) for  $\alpha \geq 0.7$ , when the alternative is Block-Basu bivariate distribution.

## References

1. BARLOW, R.E. & PROSCHAN. F. (1981). *Statistical Theory of Reliability and Life Testing*. Silver Spring, Maryland: Holt, Rinehart & Winston.
2. BLOCK, H.W. & BASU, A.P. (1974). A continuous bivariate exponential extension. *J. Amer. Statist. Assoc.*69, 1031-1037.
3. CAMBANIS. S. SIMONS. G. & STOUT. W. (1976). Inequalities for  $E k(X,Y)$  when the marginals are fixed. *Z. Wahrsch. Verw. Gebiete* 36, 285-294.
4. JOAG-DEV, K. (1984). Measures of dependence. In *Handbook of Statistics Vol. 4, Nonparametric Methods*, eds. P.R. Krishnaiah and P.K. Sen, 79-88. Newyork: North Holland.
5. KOCHAR, S.C. & GUPTA. R. P. (1987). Competitors of the Kendall-tau test for testing independence against positive quadrant dependence, *Biometrika* 74, 664-666.
6. LEHMANN E.L. (1966). Some concepts of dependence. *Ann.Math Statist.*37, 1137-1153.
7. MORGENSTERN. D. (1956). Einfache Bieispiele zweideimensionaler Verteilungen. *Mitt. Math. Statist.*8, 234-235.
8. PURI M.L. & SEN. P.K. (1971). *Nonparametric methods in Multivariate Analysis*. Newyork: Wiley.
9. SCHRIEVER.B.F. (1987). An ordering for positive dependence. *Ann. Statist.*15, 1208-1214.
10. SHAKED M. (1977). A family of concepts of dependence of bivariate distributions.*J.Amer.Statist.Assoc.*72, 642-650.
11. Shetty, I .D. and Pandit, P.V. (1996). A distribution-free test for positive quadrant dependence. *Journal of Indian Society for Probability and Statistics*, Vol. 3-4, 41-52.
12. Shetty, I .D. and Pandit, P.V. (1998). A class of distribution-free test for testing independence against positive quadrant dependence. *Assam Statistical Review*, 12, (1-2), 42-50.
13. Shetty, I .D. and Pandit, P.V. (2003). Distribution-free tests for independence against positive quadrant dependence: a generalization. *Statistical Methods and Applications*, Vol.12, 5-17.
14. WOODWORTH.C.G. (1965). On the asymptotic theory of tests of independence based on bivariate layer ranks. Abstract in *Ann.Math. Statist*, 36, 1608.