Statistics and Applications {ISSN 2454-7395 (online)}
Volume 23, No. 2, 2025 (New Series), pp 213–220
http://www.ssca.org.in/journal

Nonparametric Tests Based on Ranks for Independence against Weighted Alternative with Missing Values

Parameshwar V. Pandit[1], Deepa Yogesh Kamat[2] and Shubhashree Joshi[3]
[1]*Department of Statistics, Bangalore University, Bangalore-560056, India*
[2]*Department of Statistics, Nrupathunga University, Bengaluru-560001, India*
[3]*Department of Statistics, St. Joseph's University, Bengaluru-560027, India*

Received: 17 March 2024; Revised: 12 August 2024; Accepted: 03 February 2025

Abstract

The problem of testing independence of bivariate random variables against a weighted alternative model with possible missing values on both responses is considered in this study. Shie, Bai and Tsai (2000) studied the problem of independence with a contaminated alternative. Pandit (2006) considered the problem of independence against weighted alternatives. In this paper, a generalization of Pandit (2004) is considered the alternative with weighted contamination on both the variables for testing independence against dependence. A new rank test based on ranks is proposed and its asymptotic normality is established. Also, locally most powerful test for the above stated model due to Pandit (2004) is derived when the sample has missing values. The asymptotic null distributions of the test statistics are also provided for the purpose of practical use.

Key words: Independence tests; Locally most powerful rank test; Weighted dependence alternative; Missing values.

AMS Subject Classifications: 62G10

1. Introduction

One of the earliest problems that has attracted many statisticians working in the area of nonparametric testing is the problem of independence in bivariate situations. Many researchers have attempted to quantify the concept of stochastic dependence for bivariate distributions. Researchers such as Spearman (1904), Kendall (1962), Bhuchongkul (1964), Puri and Sen (1971), Shirahata (1974) among others proposed measure of dependence in bivariate distributions to develop rank test for independence. Pearson's correlation coefficient based on Savage (1956) scores was proposed by Iman and Conover (1987) with a view to give more weight on top ranks. In testing independence, the locally most powerful rank test (LMPR) against weighted alternatives was studied by Shei, Bai and Tsai (2000) along with some new rank tests. Pandit (2006) proposed some rank tests with different weighted

Corresponding Author: Parameshwar V. Pandit
Email: panditpv12@gmail.com

alternative and LMPR was derived. In practical situations to handle missing observations the tests mentioned above cannot be applied. For this set up a locally most powerful rank test (LMPR) for independence was derived with alternative due to Hajek and Sidak (1967). Pandit and Savitha Kumari (2015) proposed some new rank tests for testing independence against the alternative due to Hajek and Sidak (1967) and Shei, Bai and Tsai (2000). They also derived LMPR test for independence against the alternative due to Shei, Bai and Tsai (2000). Some of the alternatives for independence considered in the literature are stated below. The first family which is due to Hajek and Sidak (1967) is given by

(1) $X = X^* + \Delta Z, Y = Y^* + \Delta Z$ where X^*, Y^* and Z are mutually independent.

A similar model due to Shei, Bai and Tsai (2000) with some weights on the first component is

(2) $X = X^* + u(X^*)\Delta Z, Y = Y^* + \Delta Z$ where X^*, Y^* and Z are mutually independent and $u(x)$ is monotone in x and y respectively. However, a more general model which is due to Pandit (2006) is given by

(3) $X = X^* + u(X^*)\Delta Z, Y = Y^* + w(Y^*)\Delta Z$ where X^*, Y^* and Z are mutually independent and $u(x)$,$w(y)$ are monotone in x and y respectively.

In all the models considered above, Δ is a dependence parameter and $\Delta = 0$ implies that X and Y are independent. The model (1) is a straight forward model which makes both X and Y identically distributed if X^* and Y^* are identically distributed where as in (2) and (3), even if X^* and Y^* are identically distributed, X and Y need not be identically distributed. Here, Models (2) and (3) are weighted models. However, model (3) is the most general which includes models (1) and (2) as its particular cases. If $u(x) = w(y) = 1$, then (3) reduces to (1) and if $w(y) = 1$, (3) reduces to (2). In other words, the models due to Hajek and Sidak (1967) and Shei, Bai and Tsai (2000) are particular cases of model due to Pandit (2006). Based on a random sample $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n), (X_{n+1}, .), \ldots, (X_{n+m}, .), (., Y_{n+1})$, $\ldots$, $(.\ , Y_{n+k}))$ from a bivariate distribution function F(x, y), we consider the problem of testing $H_0 : F(x, y) = F_1(x).F_2(y)$, forall (x, y) against $H_0 : F(x, y) \neq F_1(x).F_2(y)$, for some (x, y). Here, the alternative considered is a more general weighted alternative (3) due to Pandit (2006).
Under (3), it is clear that if $\Delta = 0$, X and Y are independent and larger the Δ is more dependent are X and Y. Thus, the constant Δ may be regarded as a dependence or mixing coefficient. The alternatives stated in (3) indicate the positive dependence of the random variables X and Y. If we assume negative dependence between X and Y, then the model (3) is (3*) $X = X^* + u(X^*)\Delta Z, Y = Y^* - w(Y^*)\Delta Z$ and $X = X^* - u\,(X^*)\,\Delta Z, Y = Y^* + w(Y^*)\Delta Z$ where X^*, Y^* and Z are mutually independent and $u(x)$, $w(y)$ are monotone in x and y respectively.

The first coordinates $X_1, X_2, \ldots, X_{n+m}$ of the sample having the ranks $R_1, R_2, \ldots$, $R_{n+m}$and the the second coordinates$Y_1, Y_2, \ldots, Y_{\mathrm{n+k}}$ of the sample have the ranks Q_1, Q_2, $\ldots$, Q_{n+k}. We present rank tests for testing independence against the alternative (3) due to Pandit (2006). Naturally the test statistics proposed here are the functions of R_{i}'s and Q_{j}'s, $i = 1, 2, \ldots, n+m, j = 1, 2, \ldots, n+k$.

In Section 2, a new test based on rank statistic for testing independence is proposed and the LMPR test for the alternative (3) is considered in Section 3. Power comparison is given in Section 4. Section 5 is devoted to some remarks and conclusions.

2. Testing independence based on rank statistic

The model of dependence considered is that considered in Pandit (2006). The model for a random sample with missing values is as specified below. Let

$$X_i = X_i^* + \Delta u\left(X_i^*\right) Z_i, \qquad Y_i = Y_i^* + \Delta w\left(Y_i^*\right) Z_i, \quad i = 1, 2, ..., n$$

$$X_j = X_j^* + \Delta u\left(X_j^*\right) Z_j, \quad Y_l = Y_l^* + \Delta w\left(Y_l^*\right) Z_{m+l}, \quad j = 1, 2, ..., m; l = 1, 2, ..., k.$$

The variables X^*, Y^* and Z are independent and Δ is a real nonnegative parameter.

Spearman (1904) introduced the correlation coefficient-based ranks of $(X_{\mathrm{i}}, Y_{\mathrm{i}})$, $i = 1, 2, \ldots, n$. Further, Shei, Bai and Tsai (2000) proposed a statistic based on rank correlation with weights. In this paper, the statistic is proposed by extending it to suit the above model (3). The proposed test statistic for testing bivariate independence against model (3) is given by,

$$S_P = \sum_{i=1}^{n} u_i w_i \left(R_i - \tfrac{n+m+1}{2}\right)\left(Q_i - \tfrac{n+k+1}{2}\right),$$

where, $u_i = I(i \leq {n_1}^*)$, $w_i = I(i \leq {n_2}^*)$, where ${n_1}^* = (n+1)p$, ${n_2}^* = (n+1)q$ and $0 \leq p, q \leq 1$, were p, q roughly the proportion of the observed items in X and Y respectively. The choice of p, q is to have less loss in level of significance. (Pandit and Savitha Kumari (2015)). The asymptotic distribution of the statistic S_P is given in the following theorem. Let I(f) denote the Fisher information, $I(f) = \int_{-\infty}^{\infty}\left(\frac{f'(x)}{f(x)}\right)^2 f(x)dx$

2.1. Theorem 1

Assume that H_0 holds, $I(f_{10}) < \infty$ and $I(f_{20}) < \infty$. Then, $\frac{S_p}{\sqrt{\frac{n(n^2-1)}{12}}}$ converges in distribution to normal variate with mean zero and variance σ^2 where $\sigma^2 = \begin{cases} p((1-p)^3 + p^2) & p < q \\ q((1-q)^3 + q^2) & p \geq q \end{cases}$

Proof: Here the statistic S_p is a linear rank statistic with $E(S_p) = 0$. Assuming that $I(f_{10}) < \infty$ and $I(f_{20}) < \infty$, the asymptotic variance of $\frac{S_p}{\sqrt{\frac{n(n^2-1)}{12}}}$ under H_0 is given by,

$\mathrm{Var}\left(\frac{S_p}{\sqrt{\frac{n(n^2-1)}{12}}}\right) \to \sigma^2 = \begin{cases} p((1-p)^3 + p^2) & p < q \\ q((1-q)^3 + q^2) & p \geq q \end{cases}$ as n$\to \infty$. By applying theorem 1.6a of chapter V in Hajek and Sidak (1967), $\frac{S_p}{\sqrt{\frac{n(n^2-1)}{12}}}$ converges in distribution to normal variate with mean zero and variance σ^2 where $\sigma^2 = \begin{cases} p((1-p)^3 + p^2) & p < q \\ q((1-q)^3 + q^2) & p \geq q \end{cases}$

3. Locally most powerful rank test for weighted alternative

To develop locally most powerful rank (LMPR) test we consider the model due to Pandit (2006). For that, X^* and Y^* are the absolutely continuous random variables with probability density functions (p.d.f) $f_{10}(x)$ and $f_{20}(y)$ respectively and the distribution of Z is arbitrary. For given $(x, (\Delta)z)$ the equation $x = x^{*}+w(x^{*})\Delta z$ has the unique solution which is given by X^{*}=t (x, Δz) and for (y, Δz) the equation y=Y^{*}+u(Y^{*})Δz has the unique solution which is given by Y^{*}=t (y, Δz). Then the Jacobian is

$$|J| = \left|\frac{\partial(x^{*}, y^{*}, z)}{\partial(x, y, z)}\right| = \{(1 + \Delta z u(x^{*})(1 + \Delta z w(y^{*})\}^{-1}$$

The joint p.d.f., g_{Δ}, of $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n), (X_{n+1}, .), \ldots, (X_{n+m}), (., Y_{n+1}), \ldots,$ $(., Y_{n+k})$ is given by

$$g_{\Delta} = \prod_{i=1}^{n} h_{\Delta}(x_i, y_i). \prod_{j=n+1}^{n+m} f_{10}(x_j^{*}). \prod_{l=n+1}^{n+k} f_{20}(y_l^{*})$$

where $h_{\Delta}(x, y) = \int_{-\infty}^{\infty} f_{10}(x^{*}).f_{20}(y - \Delta z).dM(z)$ and $M(z)$ is the distribution function of Z with mean μ_z and finite variance σ_z^2. Let $X_{(i)}$ and $Y_{(i)}$ be the i-th order statistic of $\{X_1, X_2, \ldots, X_{n+m}\}$ and $\{Y_1, Y_2, \ldots, Y_{n+k}\}$, respectively.

Further, let $a_{n+m}(r_i, f_{10}) = E\left\{-\left(\frac{(uf_{10})'}{f_{10}}\right)(X_{(i)})\right\}$ and $a_{n+k}(q_i, f_{20}) = E\left\{-\left(\frac{(wf_{20})'}{f_{20}}\right)(Y_{(i)})\right\}$ denote the score functions corresponding to f_{10} and f_{20} respectively. In order to obtain LMPR test we assume the following conditions:

(i) The derivatives $(uf_{10})'$ and $(wf_{20})'$are continuous,

(ii) $\int_{-\infty}^{\infty} \left|(uf_{10})'(x)\right| dx < \infty$ and $\int_{-\infty}^{\infty} \left|(wf_{20})'(x)\right| dx < \infty$,

(iii) The missing observations on either the first coordinate or the second coordinate occur at random.

The following theorem states the LMPR test.

3.1. Theorem 2

Under the conditions (i) to (iii) and for the model (1), the test statistic $D_1 = \sum_{i=1}^{n} a_{n+m}(r_i, u, f_{10}).a_{n+k}(q_i, w, f_{20})$ with critical region $D_1 > c$, where c is a constant, is locally most powerful rank test for testing $H_0 : \Delta = 0$ against $H_1 : \Delta > 0$ at corresponding level of significance.

Corollary 1: If f_{10} and f_{20} are from Logistic family, then the test based on S_P with critical region $S_P \geq c$,where c is a constant, is asymptotic LMPR test for testing $H_0 : \Delta = 0$ against $H_1 : \Delta > 0$ for model (3).

4. Power comparison

In this Section, we compare the powers of the proposed test S_P with those of top-down statistic r_T and Kendall's τ. The alternative used is as in (1). For power comparisons, the alternatives considered are

1. X_{i}, Y_{i} and Z_{i} follow normal with mean zero and variance one.
2. X_{i}, Y_{i} and Z_{i} follow logistic (0, 1).

The correlation coefficient between X and Y denoted by ρ in terms of Δ is

$$\rho = \frac{pq\Delta^2\sigma_z^2}{\sqrt{(\sigma_x^2 + p\Delta^2\sigma_z^2)(\sigma_y^2 + q\Delta^2\sigma_z^2)}}$$

Here it is to be noted that $\rho = 0$ implies the independence. For simulation selected values of ρ are considered when $p = 0.9$.

Table 1: Empirical powers of new test S_P top-down statistic r_T and Kendall's τ for $n = 20$

m	k	ρ	$N(0,1)$			$Logistic(0,1)$		
			S_p	r_T	τ	S_p	r_T	τ
2	2	0.1	0.1325	0.0994	0.0812	0.1625	0.0872	0.0905
		0.4	0.3242	0.1089	0.1582	0.3255	0.0951	0.1567
		0.7	0.5122	0.1279	0.2665	0.5441	0.1056	0.2381
2	3	0.1	0.2624	0.1552	0.1652	0.2847	0.2311	0.2153
		0.4	0.6234	0.557	0.5284	0.7995	0.6732	0.5932
		0.7	0.9132	0.9081	0.8973	0.9653	0.8973	0.8693
3	3	0.1	0.2925	0.1072	0.0929	0.317	0.1991	0.2125
		0.4	0.6356	0.1187	0.1696	0.717	0.5862	0.5685
		0.7	0.9212	0.1286	0.2765	0.9664	0.8984	0.8791
3	4	0.1	0.3514	0.1645	0.1673	0.3558	0.2732	0.2053
		0.4	0.6531	0.5483	0.5096	0.8019	0.6578	0.5937
		0.7	0.9437	0.9104	0.8973	0.9762	0.9073	0.8774

Table 2: Empirical powers of new test S_P top-down statistic r_T and Kendall's τ for $n = 50$

m	k	ρ	$N(0,1)$			$Logistic(0,1)$		
			S_p	r_T	τ	S_p	r_T	τ
2	2	0.1	0.1365	0.1024	0.0836	0.1674	0.0898	0.0932
		0.4	0.3339	0.1122	0.1629	0.3352	0.098	0.1614
		0.7	0.5276	0.1317	0.2745	0.5605	0.1088	0.2452
2	3	0.1	0.2703	0.1599	0.1702	0.2932	0.238	0.2218
		0.4	0.6421	0.5737	0.5443	0.8235	0.6934	0.611
		0.7	0.9406	0.9353	0.9242	0.9943	0.9242	0.8954
3	3	0.1	0.3013	0.1104	0.0957	0.3265	0.2051	0.2189
		0.4	0.6547	0.1223	0.1747	0.7385	0.6038	0.5856
		0.7	0.9488	0.1325	0.2848	0.9954	0.9254	0.9055
3	4	0.1	0.3619	0.1694	0.1723	0.3664	0.2814	0.2115
		0.4	0.6727	0.5647	0.5249	0.8259	0.6775	0.6115
		0.7	0.972	0.9377	0.9242	1	0.9345	0.9037

Table 3: Empirical powers of new test S_P top-down statistic r_T and Kendall's τ for $n = 100$

m	k	ρ	$N(0,1)$			$Logistic(0,1)$		
			S_P	r_T	τ	S_P	r_T	τ
2	2	0.1	0.1391	0.1044	0.0853	0.4365	0.0916	0.0950
		0.4	0.3404	0.1143	0.1661	0.8221	0.0999	0.1645
		0.7	0.5378	0.1343	0.2798	1.0000	0.1109	0.2500
2	3	0.1	0.2755	0.1630	0.1735	0.5345	0.2427	0.2261
		0.4	0.6546	0.5849	0.5548	0.8751	0.7069	0.6229
		0.7	0.9589	0.9535	0.9422	1.0000	0.9422	0.9128
3	3	0.1	0.3071	0.1126	0.0975	0.6674	0.2091	0.2231
		0.4	0.6674	0.1246	0.1781	0.9843	0.6155	0.5969
		0.7	0.9673	0.1350	0.2903	1.0000	0.9433	0.9231
3	4	0.1	0.369	0.1727	0.1757	0.7203	0.2869	0.2156
		0.4	0.6858	0.5757	0.5351	0.9546	0.6907	0.6234
		0.7	0.9909	0.9559	0.9422	1.0000	0.9527	0.9213

In the Tables 1-4, m is the number of missing observations corresponding to x-values and k is the number of missing observations corresponding to y-values. From the above tables it is easily seen that the test S_P is more powerful than the top-down statistic r_T due to Iman and Conover (1987) and Kendall's test τ.

5. Remarks and conclusion

1. The problem of independence against a weighted alternative (3) when the data have missing values is studied. Pandit (2006) considered the alternative (3) which is the generalization of Hajek and Sidak (1967), accommodating weighted contamination in both coordinates of the pair.

2. A statistic, S_P is proposed for the above stated problem which is the generalization of Spearman's rank correlation coefficient accommodating missing values. The asymptotic normality of the test statistic S_P is established.

3. Locally most powerful rank (LMPR) test for testing independence against the alternative due to Pandit (2006) accommodating missing values is derived.

4. The LMPR for the alternative (1) due to Hajek and Sidak (1967) considered in Wei (1983), and that for the alternative due to Shai, Bai and Tsai (2000) established in Pandit and Savitha Kumari (2015) can be obtained as particular cases of the LMPR derived here.

5. The small sample powers of tests based on S_P, r_T , τ are evaluated using simulation study. It is observed that the power of the test based on S_P is more as compared to the tests due to Iman and Conover (1987) r_T and Kendall's test τ, when the marginal distributions of $X*$ and $Y*$ belong to normal and logistic families.

6. The test based on S_P proposed here is shown to be LMPR when the marginal distributions of $X*$ and $Y*$ belong to logistic family. It is also observed that test based on

Table 4: Empirical powers of new test S_P top-down statistic r_T and Kendall's τ for $n = 200$

m	k	ρ	$N(0,1)$			$Logistic(0,1)$		
			S_P	r_T	τ	S_P	r_T	τ
2	2	0.1	0.1405	0.1054	0.0861	0.4496	0.0924	0.0959
		0.4	0.3437	0.1154	0.1677	0.8468	0.1008	0.1661
		0.7	0.5429	0.1356	0.2825	1.0000	0.1119	0.2524
2	3	0.1	0.2781	0.1645	0.1751	0.5505	0.2450	0.2282
		0.4	0.6608	0.5904	0.5601	0.9014	0.7136	0.6288
		0.7	0.968	0.9626	0.9511	1.0000	0.9511	0.9215
3	3	0.1	0.3101	0.1136	0.0985	0.6874	0.2110	0.2253
		0.4	0.6737	0.1258	0.1798	0.9765	0.6214	0.6026
		0.7	0.9765	0.1363	0.2931	1.0000	0.9523	0.9318
3	4	0.1	0.3725	0.1744	0.1773	0.7419	0.2896	0.2176
		0.4	0.6923	0.5812	0.5402	0.9832	0.6973	0.6293
		0.7	0.9956	0.9650	0.9511	1.0000	0.9617	0.9300

S_P has better performance in terms of small sample power.

Acknowledgements

The authors are indeed grateful to the Editors for their guidance and counsel. The authors also express their gratefulness to the reviewer for valuable comments and suggestions of generously listing many useful references.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests.* Academic Press; New York.

Iman, R. L. and Conover, W. J. (1987). A measure of top-down correlation. *Technometrics*, **29**, 351-357.

Kendall, M. G. (1962). *Rank Correlation Methods. Griffin: London.*

Pandit, P. V. (2006). Locally most powerful and other rank tests for independence-with a contaminated weighted alternative. *Metrika*, **64**, 379-387.

Pandit, P. V. and Savitha Kumari (2015). Rank tests for independence against weighted alternatives with missing values. *American Journal of Mathematics and Statistics*, **5**, 196-199.

Puri, M. L. and Sen, P. K. (1971). *Non-parametric Methods in Multivariate Analysis.* Wiley: New York.

Savage, I. R. (1956). Contributions to the theory of rank order statistics- two-sample case. *Annals of Mathetmatical Statististics*, **27**, 590-615.

Shieh, G. S. (1998). A weighted Kendal's statistic. *Statistics and Probability Letters*, **39**, 17-24.

Shei, G. S., Bai, Z., and Tsai, W. Y. (2000). Rank tests for independence – with a weighted contamination alternative. *Statistica Sinica*, **10**, 577-593.

Shirahata, S. (1974). Locally most powerful rank tests for independence. *Bulletin of Mathematical Sciences*, **16**, 11-21.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychiatry*, **15**, 72-101.

Wei, L. J. (1983). Tests for independence in the presence of missing values. *Austrian Journal of Statistics*, **25**, 85-90.