March-April 2020
ISSN: 0193-4120 Page No. 1243 - 1248

Nonparametric Tests for Point of Symmetry Based on Sub Sample Extremes

[1]Savitha Kumari, [2*]Parameshwar V. Pandit

[1]Department of Statistics, SDM College (Autonomous), Ujire, D. K., India
[2]Department of Statistics, Bangalore University, Bangalore, India
[1]savi_rrao@yahoo.co.in, [2]panditpv12@gmail.com

Article Info
Volume 83
Page Number: 1243 - 1248
Publication Issue:
March - April 2020

Article History
***Article Received**: 24 July 2019*
***Revised**: 12 September 2019*
***Accepted**: 15 February 2020*
***Publication**: 14 March 2020*

Abstract

One sample location problem is widely studied problems in Nonparametric set up when the underlying distribution of the sample is symmetric. In this paper, testing for point of symmetry in one sample setting is considered by proposing a class of tests. The proposed class of tests is based on convex combination of two statistics, which are computed using subsamples. The Pitman asymptotic relative efficiencies of few members of the class are computed with respect to the other tests exist in the literature to evaluate its performance. The choice of subsample size is determined for different distributions to make the statistics easily usable.

***Keywords:** Asymptotic relative efficiency, one sample location problem, symmetry, U- statistic*

1. Introduction

One of the prominent problems studied in the area of nonparametric inference is the problem of testing for the point of symmetry in one sample setting. There are many distribution free tests available in the literature for the above problem. Two well known tests available in the literature for this problem are sign and Wilcoxon-signed rank tests. A test statistic was proposed by Madhav Rao [5] which is based on subsample median. A test procedure based on a U-statistics was proposed by Mehra, Prasad and Rao [6]whose kernel depends on a positive constant ‘a’ and subsamples of size two andAhmad [1] proposed class of Mann-Whitney-Wilcoxon type statistics for one sample location problem. Shetty and Pandit [9] generalised the test due to Mehra, Prasad and Rao [6]. Bandyopadhyay and Datta [2]studied a test procedure for this problem by proposing adaptive nonparametric tests. One sample location test for multilevel data was studied by Larocquea, Nevalainenb and Oja [3]. Rattihalli and Raghunath [8]extended the test due to Shetty and Pandit [9]which is based on subsample order statistics and medians.

In this paper, a new class of test procedures based on convex combination of U-statistics is proposed, in which one of them being the statistic due to Ahmad[1]. The mean and asymptotic variance of the proposed class of test statistics are derived for arbitrary k, the sub sample size. The Pitman asymptotic relative efficiencies relative to t-test are computed to evaluate the performance of few members of the proposed class.

The organization of the paper is as below. In section 2, new classes of test statistics are proposed. The distributional properties of proposed class of test based on $\mathbf{U}_{\delta}$ **(k)** are studied in section 3. Section 4 contains the asymptotic relative efficiency comparison. Asymptotic distribution of test statistic U (k) is considered in Section 5 along with its performance in asymptotic sense.Some concluding remarks are presented in Section 6.

2. Proposed Test Procedure

Let $(X_1, X_2,...X_n)$ be a random sample of size n from a population with distribution function F(x-θ), where F is absolutely continuous symmetric distribution function. Here θ is the point of symmetry. The problem considered here is testing $H_0 : \theta = 0 \ Vs \ H_1 : \theta \neq 0.$

Let $h_1(X_1, X_2,...X_{k+2}) = \begin{cases} 1 & if\ Min(X_1,...X_k) > -Max(X_{k+1}, X_{k+2}) \\ 0 & Otherwise \end{cases}$

and $h_2(X_1, X_2,...X_{k+1}) = \begin{cases} 1 & if\ Min(X_1,...X_k) > -X_{k+1} \\ 0 & Otherwise \end{cases}$

Then the U-statistic with kernel h_1 is given by

$$U_1(X_1, X_2,...X_n) = \frac{1}{\binom{n}{k+2}\binom{k+2}{2}} \sum_{C_1} h_1(X_{i_1}, X_{i_2},...X_{i_{k+2}})$$

Where $\sum_{C_1}$ is the summation over all $\binom{n}{k+2}$ combinations of possible integers $\{i_1, i_2,..., i_{k+2}\}$ taken out of the set $\{1,2,..., n\}$ without replacement.

Also, the U-statistic based on h_2

$$U_2(X_1, X_2,...X_n) = \frac{1}{\binom{n}{k+1}\binom{k+1}{1}} \sum_{C_2} h_2(X_{i_1}, X_{i_2},...X_{i_{k+1}})$$

Where $\sum_{C_2}$ is summation over all $\binom{n}{k+1}$ combinations of possible integers $\{i_1, i_2,..., i_{k+1}\}$ taken out of the set $\{1,2,..., n\}$ without replacement.

For testing H_0 against H_1, we propose

(i) A convex combination of above U-statistics given by,

$$U_\delta(k) = \delta U_1(k) + (1-\delta) U_2(k)\ , \qquad 0 \le \delta \le 1$$

(ii) A linear combination of above U-statistics given by,

$$U\ (k) = U_1(k) - U_2(k)$$

3. Asymptotic distribution of test statistic $U_\delta(k)$

The mean of $\mathbf{U_\delta}$ **(k)** is given by

$$\gamma(F) = E[U_\delta(k)]$$
$$= \delta\, E[U_1(k)] + (1-\delta)\ E[U_2(k)]$$
$$= \delta\, P[Min(X_1, X_2,...X_k) > -Max(X_{k+1}, X_{k+2})] + (1-\delta)\, P[Min(X_1, X_2,...X_k) > -(X_{k+1})]$$
$$= \delta \int_{-\infty}^{\infty} \bar{F}^k(-x-2\theta)\, dF^2(x-\theta) + (1-\delta) \int_{-\infty}^{\infty} \bar{F}^k(-x-2\theta)\, dF(x-\theta)$$

Under H_0 , $\quad \gamma_0(F) = \dfrac{2\delta}{k+2} + \dfrac{1-\delta}{k+1}$.

Here we have $\lim_{n\to\infty} Cov(U_1, U_2) = (k+2)(k+1)\,\sigma_{1,1}$, where

$$\sigma_{1,1} = Cov[h_1, h_2]$$
$$= Cov[h_1(X_1, X_2,...X_{k+2})\, h_2(X_1, X_{k+3},...X_{2k+2})]$$
$$= E[h_1(X_1, X_2,...X_{k+2})\, h_2(X_1, X_{k+3},...X_{2k+2})] - \left(\frac{2}{k+2}\right)\left(\frac{1}{k+1}\right).$$

Now

March-April 2020
ISSN: 0193-4120 Page No. 1243 - 1248

$$E[h_1(X_1, X_2,...X_{k+2})|X_1 = x] = \frac{2}{k+2} + \frac{2k}{(k+2)(k+1)}[F^{k+1}(x) - \overline{F}^{k+1}(x)] .$$

Also

$$E[h_2(X_1, X_{k+3},...X_{2k+2})|X_1 = x] = \frac{1}{(k+1)}[1 - \overline{F}^{k}(x) + F^{k}(x)]$$

Thus, $\sigma_{1,1} = \frac{2k}{(k+1)^2\,(k+2)}\left[\frac{1}{k+1} - 2\,\beta\,(k+1, k+2)\right]$

and $\lim\limits_{n\to\infty} Cov(U_1, U_2) = (k+2)(k+1)\sigma_{1,1}$

$$= \frac{2k}{(k+1)^2}\left[1 - 2\,(k+1)\,\beta\,(k+1, k+2)\right]$$

Hence variance of $\sqrt{n}[U_\delta(k) - E(U_\delta(k))]$ under H_0 is given by

$$\sigma^2_{\delta 0}(k) \;= \delta^2\sigma^2_{10} + (1-\delta)^2\,\sigma^2_{20} + 2\,\delta\,(1-\delta)\,(k+2)\,(k+1)\sigma_{1,1}$$

Theorem 3.1: The asymptotic null distribution of $\sqrt{n}[U_\delta(k) - E(U_\delta(k))]$ as n→∞ is normal with mean zero and variance $\sigma^2_{\delta 0}(k)$. ie,

$$\sigma^2_{\delta 0}(k) = \delta^2\,\frac{8k^2}{(k+1)^2}\left\{\frac{1}{2k+3} - \beta(k+2, k+2)\right\} + (1-\delta)^2\,2\left\{\frac{1}{2k+1} - \beta(k+1, k+1)\right\}$$
$$+ 2\,\delta\,(1-\delta)\,\frac{2k}{(k+1)^2}\left[1 - 2\,(k+1)\,\beta\,(k+1, k+2)\right]$$

Variances for various values of k and δ are given below

Δ	k=2	k=3	k=4	k=5	k=6	k=7	k=8
0	0.3333	0.2714	0.2190	0.1811	0.1537	0.1333	0.1176
0.1	0.3768	0.3147	0.2588	0.2171	0.1863	0.1630	0.1448
0.2	0.4140	0.3533	0.2952	0.2506	0.2170	0.1911	0.1707
0.3	0.4448	0.3872	0.3282	0.2815	0.2457	0.2177	0.1954
0.4	0.4692	0.4164	0.3578	0.3099	0.2725	0.2427	0.2188
0.5	0.4873	0.4409	0.3839	0.3358	0.2973	0.2662	0.2410
0.6	0.4990	0.4607	0.4067	0.3591	0.3201	0.2881	0.2618
0.7	0.5044	0.4758	0.4260	0.3798	0.3409	0.3085	0.2815
0.8	0.5035	0.4862	0.4420	0.3981	0.3598	0.3273	0.2998
0.9	0.4962	0.4919	0.4545	0.4137	0.3768	0.3446	0.3169
1	0.4825	0.4929	0.4636	0.4269	0.3917	0.3603	0.3327

4. Asymptotic Relative Efficiencies of $U_\delta(k)$.

The asymptotic relative efficiency of $U_\delta(k)$ with respect to any other test T is given by,

$$ARE[U_\delta(k), T] = \left[\frac{Eff\,[U_\delta(k)]}{Eff\,(T)}\right]^2 .$$

The expression Eff $[U_\delta(k)]$ itself becomes the ARE[$U_\delta(k)$,T], when $\sigma^2 = 1$.

Here the performance of the proposed test relative toSign test(S), Wilcoxon Signed Rank test(W), test due to Mehra et.al.(1990) (T_{a*}), test due to Shetty and Pandit (2000)(U_{a*} (4,2)) and test due to Rattihalli and Raghunath(2012) (V_{a*}(3,1)) for comparison.
The various values of Eff(S), Eff(W) , Eff(T_{a*}), Eff(U_{a*} (4,2)) and Eff (V_{a*} (3,1)) for various standard distributions are given in Table 1. ARE's of $U_\delta(k)$ for

TEST Engineering & Management

March-April 2020
ISSN: 0193-4120 Page No. 1243 - 1248

various values of δ and k are given in Table 2 to Table 5.

Table 1: The AREs of different tests relative to T-test

Density	Eff (S)	Eff (W)	Eff(T_{a*})	Eff(U_{a*}(4,2))	Eff(V_{a*}(3,1))
Cauchy	0.4053	0.3040	0.4053	0.4252	0.6411
Laplace	2.0000	1.5000	2.0000	2.0000	2.0000
Logistic	0.8225	1.0966	1.0966	1.1364	1.0799
Normal	0.6366	0.9549	0.9643	0.9503	0.9868
Triangular	0.6667	0.8889	0.8889	0.7965	0.9657
Parabolic	0.4500	0.8460	0.9360	0.8125	1.0829
Uniform	0.3333	1.3333	1.3333	0.6869	2.0021
No name	0.0000	2.6667	10.6600	16.4563	14.3957

Here the no name distribution is given by the pdf $f(x) = \frac{1}{2}|x| I_{(|x| \le \sqrt{2})}$.

Table 2: ARE's $U_{\delta}(k)$ of relative to T-test for various values of δ when k=2

Distribution\ δ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Cauchy	0.3034	0.2755	0.2573	0.2457	0.2388	0.2358	0.2360	0.2392	0.2454	0.2550	0.2685
Laplace	0.3749	0.3428	0.3223	0.3098	0.3031	0.3010	0.3031	0.3090	0.3189	0.3332	0.3527
Logistic	1.0982	1.0106	0.9563	0.9246	0.9099	0.9089	0.9201	0.9431	0.9782	1.0271	1.0924
Normal	0.9551	0.8818	0.8369	0.8116	0.8010	0.8023	0.8143	0.8368	0.8701	0.9158	0.9762
Triangular	5.3336	4.9360	4.6956	4.5636	4.5134	4.5297	4.6068	4.7426	4.9402	5.2085	5.5614
Parabolic	0.8639	0.8037	0.7684	0.7504	0.7455	0.7515	0.7675	0.7933	0.8295	0.8778	0.9407
Uniform	0.7501	0.9031	1.0735	1.2646	1.4800	1.7243	2.0039	2.3266	2.7032	3.1488	3.6843
No name	2.6669	2.5917	2.5806	2.6176	2.6947	2.8084	2.9596	3.1509	3.3880	3.6810	4.0442

Table 3: ARE's $U_{\delta}(k)$ of relative to T-test for various values of δ when k=4

Distribution\ δ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Cauchy	0.2249	0.2023	0.1880	0.1790	0.1736	0.1707	0.1698	0.1706	0.1728	0.1765	0.1814
Laplace	0.3208	0.2925	0.2754	0.2654	0.2603	0.2588	0.2601	0.2638	0.2697	0.2777	0.2878
Logistic	1.0692	0.9842	0.9351	0.9087	0.8979	0.8991	0.9095	0.9284	0.9547	0.9886	1.0299
Normal	0.9877	0.9136	0.8719	0.8507	0.8437	0.8478	0.8604	0.8810	0.9085	0.9432	0.9851
Triangular	5.7642	5.3520	5.1250	5.0163	4.9900	5.0278	5.1156	5.2500	5.4257	5.6445	5.9067
Parabolic	1.0250	0.9580	0.9229	0.9083	0.9081	0.9192	0.9393	0.9679	1.0039	1.0480	1.1003
Uniform	1.1410	1.4378	1.7561	2.0990	2.4695	2.8718	3.3086	3.7862	4.3086	4.8847	5.5220
No name	5.2990	4.2679	3.5551	3.0342	2.6372	2.3255	2.0735	1.8669	1.6938	1.5478	1.4229

Table 4: ARE's $U_{\delta}(k)$ of relative to T-test for various values of δ when k=6

Distribution\ δ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Cauchy	0.1464	0.1315	0.1224	0.1169	0.1137	0.1120	0.1116	0.1121	0.1134	0.1154	0.1180
Laplace	0.2571	0.2344	0.2213	0.2141	0.2105	0.2097	0.2111	0.2139	0.2183	0.2239	0.2307
Logistic	0.9845	0.9050	0.8608	0.8382	0.8294	0.8308	0.8405	0.8560	0.8772	0.9035	0.9345
Normal	0.9709	0.8965	0.8563	0.8368	0.8308	0.8349	0.8469	0.8648	0.8883	0.9169	0.9503

March-April 2020
ISSN: 0193-4120 Page No. 1243 - 1248

Triangular	5.9961	5.5624	5.3350	5.2333	5.2135	5.2548	5.3455	5.4725	5.6346	5.8290	6.0531
Parabolic	1.1625	1.0840	1.0445	1.0288	1.0287	1.0404	1.0616	1.0898	1.1250	1.1666	1.2141
Uniform	1.6265	2.0729	2.5430	3.0403	3.5657	4.1230	4.7169	5.3462	6.0176	6.7354	7.5027
No name	8.9887	8.5899	8.4567	8.4899	8.6345	8.8662	9.1718	9.5340	9.9537	10.4290	10.9572

Table 5: ARE's $U_{\delta}(k)$ of relative to T-test for various values of δ when k=8

Distribution\ δ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Cauchy	0.0955	0.0857	0.0799	0.0764	0.0744	0.0734	0.0732	0.0735	0.0743	0.0754	0.0769
Laplace	0.2083	0.1902	0.1802	0.1748	0.1725	0.1722	0.1735	0.1760	0.1796	0.1840	0.1892
Logistic	0.8817	0.8106	0.7728	0.7539	0.7474	0.7495	0.7584	0.7721	0.7904	0.8124	0.8379
Normal	0.9260	0.8543	0.8169	0.7990	0.7941	0.7980	0.8092	0.8252	0.8462	0.8710	0.8997
Triangular	6.0903	5.6474	5.4246	5.3273	5.3137	5.3571	5.4477	5.5704	5.7261	5.9072	6.1138
Parabolic	1.0345	0.9945	0.9855	0.9946	1.0161	1.0465	1.0846	1.1282	1.1779	1.2323	1.2919
Uniform	2.1259	2.7217	3.3443	3.9933	4.6726	5.3821	6.1286	6.9077	7.7292	8.5912	9.4998
No name	12.9528	12.2863	12.0369	12.0275	12.1824	12.4510	12.8178	13.2525	13.7601	14.3258	14.9511

5. Asymptotic comparison of test statistic U(k)

The expectation of **U (k)** is given by

$$\gamma(F) = E[U\ \ (k)]$$

$$= E[U_1(k)] - E[U_2(k)]$$

$$= P[Min(X_1, X_2, ... X_k) > -Max(X_{k+1}, X_{k+2})] - P[Min(X_1, X_2, ... X_k) > -(X_{k+1})]$$

$$= \int_{-\infty}^{\infty} \bar{F}^{k}(-x-2\theta)\,dF^2(x-\theta) - \int_{-\infty}^{\infty} \bar{F}^{k}(-x-2\theta)\,dF(x-\theta)$$

Under H_0,

$$\gamma_0(F) = \frac{k}{(k+2)(k+1)}$$

The asymptotic variance of $\sqrt{n}[U(k) - E(U(k))]$ under H_0 is given by

$$\sigma_0^2(k) = \sigma_{10}^2 + \sigma_{20}^2 - 2\,(k+2)(k+1)\sigma_{1,1}$$

Variances for various values of k are given below

k	2	3	4	5	6	7	8
variance	0.0159	0.0363	0.0487	0.054	0.0554	0.0556	0.0543

Table 6 gives the AREs of U(k) relative to t-test.

Table 6: AREs of $U(k)$ relative to T-test for various distributions

Distribution	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Cauchy	0.1104	0.1059	0.0949	0.0856	0.0763	0.0647	0.0537
Laplace	0.2189	0.2153	0.2062	0.1968	0.1871	0.1745	0.1641
Logistic	0.9208	0.9125	0.8799	0.8640	0.8422	0.8007	0.7815
Normal	0.9376	0.9236	0.9107	0.9086	0.9041	0.8832	0.8680
Triangular	5.8430	5.7559	5.7898	5.9262	6.0680	6.1110	6.1931
Parabolic	1.1822	1.1645	1.1857	1.2336	1.2858	1.3160	1.7334

Uniform	43.6723	27.5482	24.8460	25.2072	26.6160	28.1025	30.0695
No name	12.9690	10.1200	11.0839	12.6562	14.5010	16.2799	18.2728

6. Some Remarks and Conclusions

1. For one-sample location problem, a class of test statistics is proposed in this paper assuming that the underlying distribution of the sample drawn is symmetric. Here, the test proposed in this situation is testing for point of symmetry in one-sample problem.

2. The asymptotic variance of the few members, $U_{\delta}(k)$ (for k=2,3,4,5,6,7,8) of the class of test statistics for various values of δ are computed as a ready reference for the researchers.

3. From tables 1 to 5, it is observed that the proposed test $U_{\delta}(k)$ performs better than the tests existing in the literature for this problem for the Triangular, Parabolic, Uniform and No name distributions. Similar performance is observed for U(k).

4. For Cauchy distribution, test due to Rattihalli and Raghunath(2012) is better.

5. The AREs of the proposed test decreases with k (that is with subsample size) for heavy tailed distributions whereas for light tailed distributions, the ARE increases with subsample size.

References

[1] Ahmad,I.A., "A class of Mann-Whitney-Wilcoxon type statistics", The American Statistician, Vol.50, No.4, (1996), 324-327.

[2] Bandyopadhyay,U. and Datta,D., "Adaptive Nonparametric tests for Single sample location problem", Statistical Methodology, 4, (2007). 424-433.

[3] Larocquea,D., Nevalainenb,J. and Oja,H., "One –sample location tests for multilevel data", Journal of Statistical Planning and Inference, 138, (2008). 2469-2482.

[4] Lehmann,E.L., and D'Abrera, H.J.M. "Nonparametrics: Statistical Methods based on Ranks", Sringer-Verlag, New York, (2006)

[5] Madhava Rao,K.S., "A one-sample test based on subsamples medians. Communications in Statistics-Theory and Methods" 19, (1990), 4559-4568.

[6] Mehra, K.L., Prasad, N.G. N. and Mahava Rao, K.S.,"A class of Nonparametric Tests for the One-Sample Location Problem" ,Australian Journal of Statistics 32,(1990),373-392.

[7] Randles, R.H. and Wolfe,D.A., "Introduction to the Theory of Nonparametric Statistics", John Wiley and Sons, New York, (1979).

[8] Rattihalli, R. N. and Raghunath,M., "Generalized nonparametric tests for one-sample location problem based on sub-samples", ProbStatForum, 5, (2012). 112-123.

[9] Shetty,I.D. and Pandit, P.V., "A Class of Distribution-free Tests for One-Sample Location Problem", Journal of the Karnatak University-Science, 43, (2000), 36-41.

[10] Stephenson, W.R., "A general class of one-sample nonparametric teststatistics based on subsamples", J.Amer.Statist.Assoc.,76, (1981), 960-966.