ProbStat Forum, Volume 05, October 2012, Pages 112–123
ISSN 0974-3235

ProbStat Forum is an e-journal.
For details please visit www.probstat.org.in

Generalized nonparametric tests for one-sample location problem based on sub-samples

Rattihalli R N[a], Raghunath M[b]

[a]*Department of Statistics, Central University of Rajasthan, Kishangarh, India*
[b]*Department of Statistics, Shivaji University, Kolhapur, India*

Abstract. A class of distribution-free tests based on U-Statistics is proposed for the one-sample location problem. The kernel defined is based on a sub-sample of size k taken from n independent identically distributed observations. This kernel depends on a constant a and r^{th}, $(k-r+1)^{th}$ order statistics together with the median of the sub-sample. For given k and r, the optimal value of the constant a is obtained by maximizing the efficacy. The performance of the test is evaluated for various symmetric models by means of asymptotic relative efficiency relative to sign test, Wilcoxon signed-rank test and other competitors. Using extensive simulations, performances of the tests based on the empirical power is also studied for some standard distributions and for a typical class of heavy tailed distributions.

1. Introduction

One of the important fundamental problems extensively considered in the nonparametric inference is one-sample location problem. The problem is to test for the location parameter that is median of a distribution when the samples are drawn from a continuous symmetric distribution.

Let $X_1, \ldots, X_n$ be a random sample of size n from an absolutely continuous symmetric distribution with cumulative distribution function (cdf),

$$F_\theta(x) = P[X_i < x] = G(x-\theta),$$

where G admits density g satisfying $g(x) = g(-x)$ for all $-\infty < x < \infty$, θ is the location parameter and the median of the distribution F.

The problem of interest is to test the hypothesis

$$H_0 : \theta = 0 \quad vs \quad H_1 : \theta \neq 0.$$

Some of the well known nonparametric tests in the literature to test the above hypothesis are Sign and Wilcoxon signed-rank tests and their generalizations.

Madhava Rao (1990) proposed a test by using the sub-samples median. Madhava Rao (1990) proposed a class of tests T_a by using a kernel that depends on an arbitrary constant a. For different $G(\cdot)$'s they have obtained the optimal values of a by maximizing the efficacy. Recently Pandit and Math (2011) used the test

2010 *Mathematics Subject Classification.* Primary: 62G10; Secondary: 62G30, 62G99
Keywords. Nonparametric tests, asymptotic relative efficiency, U-Statistics, sign test, Wilcoxon signed-rank test
Received: 26 August 2010; Revised: 26 April 2012; Accepted: 15 September 2012
Email addresses: `rnr5@rediffmail.com` (Rattihalli R N), `raghu_mmr@yahoo.co.in` (Raghunath M)

statistic T_a to develop nonparametric control chart for location. Shetty and Pandit (2000) have generalized the test proposed by Mehra et al. (1990) by considering relative positions of two symmetric order statistics taken from a sub-sample of size k. They denoted the statistic as $U_a(k,r)$. In particular, they studied the performance (in Pitman sense) of $U_a(4,2)$ and $U_a(5,2)$ with their optimal values and have showed that the first one performs better. Bandyopadhyay and Datta (2007) have proposed an adaptive nonparametric tests for a single sample location problem. Larocquea et al. (2008) have developed one-sample location tests for multi-level data.

It is well know that sign test performs better when the underlying distributions are heavy tailed. When the tails are moderate, Wilcoxon signed-rank test performs better (see Randles and Wolfe (1979)). Even though the problem of location in univariate case seems to be pretty old, researchers are finding some scope to improve the earlier proposed tests.

In this article, we propose a class of distribution-free tests based on U-statistics, which is the modification of the test proposed by Shetty and Pandit (2000). The kernel is based on the r^{th}, $(k-r+1)^{th}$ order statistics together with the median of a sub-sample of size k taken from a random sample of size n. General expressions for expected value and asymptotic variance of the proposed test statistics are derived for an arbitrary (k,r). The optimal value of constant a is obtained for different values of (k,r) by maximizing the efficacy. Note that, the proposed class of tests includes T_a and $U_a(4,2)$ as a special cases. The performance of the test is evaluated by means of Asymptotic Relative Efficiency (ARE) relative to sign test, Wilcoxon signed-rank test and other competitors. Empirical power study is carried out for some standard symmetric distributions and for a typical class of heavy tailed distributions. Some useful results to compute expected value and asymptotic variance with an illustrative example are given in Appendix.

2. Proposed class of tests

Let $X_1, \ldots, X_n$ be a random sample from an absolutely continuous distribution function $F_\theta(x) = G(x-\theta)$, where $G(y)+G(-y)=1$. The proposed statistic to test the hypothesis $H_0: \theta = 0$ against the alternative $H_1: \theta \neq 0$ is a U-statistics,

$$V_a(k,r) = \frac{\sum_c \phi(X_1, X_2, ..., X_k)}{\binom{n}{k}}$$

where the summation is over all $\binom{n}{k}$ combinations of the integers $\{1,2,\ldots,n\}$, r is fixed such that $r < k-r+1$ and

$$\phi_a(x_1,x_2,\ldots,x_k) = \begin{cases} 1, & \text{if } x_{(r)} > 0 \\ a(-a), & \text{if } x_{(r)}x_{(k-r+1)} < 0,\ x_{(r)} + x_{(k-r+1)} > (<)0,\ Med(x_1,x_2,\ldots,x_k) > (<)0 \\ -1, & \text{if } x_{(k-r+1)} < 0 \\ 0, & \text{otherwise} \end{cases}$$

where $x_{(i)}$ is the i^{th} order statistic and $Med(x_1,x_2,\ldots,x_k)$ is the median of a sub-sample of size k. The test statistic rejects H_0 for large values of $|V_a(k,r)|$. When k is even ($k=2m$, say), then median of X_1, $X_2, \ldots, X_k$ is any number in between $X_{(m)}$ and $X_{(m+1)}$, however for definiteness one may define it to be, $(X_{(m)} + X_{(m+1)})/2$. If k is odd ($k = 2m-1$, say), then median of $X_1, X_2, \ldots, X_k$ is $X_{(m)}$. In the following, for notational simplicity we write $\phi_a(\cdot)$ as $\phi(\cdot)$.

3. The asymptotic distribution of $V_a(k,r)$

In the following we derive general expressions for expectation and asymptotic variance for $V_a(k,r)$.

3.1. Expectation of $V_a(k,r)$

To obtain $E[V_a(k,r)] = E(\phi(X_1, X_2, \ldots, X_k))$, it is enough to compute the probabilities of the following events:

$$\begin{aligned} E_1 &= \left\{X_{(r)} > 0\right\} \\ E_2 &= \left\{X_{(r)}X_{(k-r+1)} < 0, X_{(r)} + X_{(k-r+1)} > 0, Med(X_1, X_2, \ldots, X_k) > 0\right\} \\ E_3 &= \left\{X_{(r)}X_{(k-r+1)} < 0, X_{(r)} + X_{(k-r+1)} < 0, Med(X_1, X_2, \ldots, X_k) < 0\right\} \\ E_4 &= \left\{X_{(k-r+1)} < 0\right\} \end{aligned}$$

The probabilities of the above events are given by,

$$P[E_1] = \sum_{i=k-r+1}^{k} \binom{k}{i} (1 - G(-\theta))^i G^{k-i}(-\theta) = 1 - B(G(-\theta); r, k - r + 1). \tag{1}$$

$$P[E_4] = \sum_{i=0}^{r-1} \binom{k}{i} G^i(-\theta)(1 - G(-\theta))^{k-i} = B(G(-\theta); k - r + 1, r). \tag{2}$$

where $B(\cdot; p, q)$ is the beta distribution function of first kind with parameters p, q.

In further derivations we consider k is odd. We obtain $P[E_2]$ and $P[E_3]$ by considering the joint density of $X_{(r)}$, $X_{((k+1)/2)}$ and $X_{(k-r+1)}$. We have that

$$P[E_2] = \int_{-\infty}^{0} \int_{-u}^{\infty} \int_{0}^{w} f(u, v, w) dv dw du,$$

where $f(u, v, w)$ is the joint density of $X_{(r)}$, $X_{((k+1)/2)}$ and $X_{(k-r+1)}$. Further,

$$P[E_2] = C \int_{-\infty}^{0} \int_{-u}^{\infty} \int_{0}^{w} [F(u)(1 - F(w))]^{r-1} \left\{[F(v) - F(u)][F(w) - F(v)]\right\}^{\frac{(k-2r-1)}{2}} dF(v)dF(w)dF(u),$$

where $C = \frac{k!}{[(r-1)!((k-2r-1)/2)!]^2}$.

Using the relation $F_\theta(x) = G(x - \theta)$, we get,

$$\begin{aligned} P[E_2] &= C \int_{-\infty}^{0} \int_{-u}^{\infty} \int_{0}^{w} [G(u - \theta)(1 - G(w - \theta))]^{r-1} [G(v - \theta) - G(u - \theta)]^{\frac{(k-2r-1)}{2}} \\ &\quad \times [G(w - \theta) - G(v - \theta)]^{\frac{(k-2r-1)}{2}} dG(v - \theta) dG(w - \theta) dG(u - \theta). \end{aligned} \tag{3}$$

Similarly,

$$\begin{aligned} P[E_3] &= C \int_{-\infty}^{0} \int_{0}^{-u} \int_{u}^{0} [G(u - \theta)(1 - G(w - \theta))]^{r-1} [G(v - \theta) - G(u - \theta)]^{\frac{(k-2r-1)}{2}} \\ &\quad \times [G(w - \theta) - G(v - \theta)]^{\frac{(k-2r-1)}{2}} dG(v - \theta) dG(w - \theta) dG(u - \theta). \end{aligned} \tag{4}$$

From Lemma 6.1 given in Appendix, we have,

$$P_\theta[E_1] = P_{-\theta}[E_4] \quad \text{and} \quad P_\theta[E_2] = P_{-\theta}[E_3].$$

Hence,

$$\mu(\theta, k, r) = E[V_a(k, r)] = P[E_1] + a\{P[E_2] - P[E_3]\} - P[E_4]. \tag{5}$$

Note that under $H_0 : \theta = 0$, $P_0[E_1] = P_0[E_4]$ and $P_0[E_2] = P_0[E_3]$. Thus $E_{H_0}[V_a(k, r)] = 0$.

3.2. Asymptotic variance of $V_a(k, r)$

Since $V_a(k, r)$ is a one-sample U-statistics, from Randles and Wolfe (1979), the asymptotic distribution of $V_a(k, r)$ follows normal with mean 0 and variance $k^2\zeta_1(k, r)$, where

$$\zeta_1(k,r) = Cov(\phi(X_1, X_2, \ldots, X_k)\phi(X_1, X_{k+1}, \ldots, X_{2k-1})) = Var_{H_0}\left[E_{H_0}(\phi(X_1, X_2, \ldots, X_k)|X_1 = x)\right].$$

To obtain the general expression of $Var\left[E_{H_0}(\phi(X_1, X_2, \ldots, X_k)|X_1 = x)\right]$, we consider the following cases.

Case I (When $x > 0$): Let $(X_1, X_2, \ldots, X_k)$ be denoted by $(X_1, Y_1, \ldots, Y_{k-1}) = (X_1, \underline{Y})$, where $\underline{Y} = (Y_1, \ldots, Y_{k-1})$. Let $y_{(1)} < y_{(2)} <, \ldots, < y_{(k-1)}$ be the ordered values of $y \in R^{k-1}$. For given x, $\phi(x, y)$ takes the respective values 1, a, $-a$, -1 and 0 on the sets,

$$\begin{aligned}
E_1(x) &= \left\{y \in R^{k-1} : y_{(r)} > 0\right\} \\
E_2(x) &= \left\{y \in R^{k-1} : \{y_m > 0\}, \text{ and } \left\{0 < y_{(k-r+1)} < x, y_{(r)} > -x \text{ or } y_{(k-r+1)} > x, y_{(r)} > -y_{(k-r+1)}\right\}\right\} \\
E_3(x) &= \left\{y \in R^{k-1} : \{y_m < 0\}, \text{ and } \left\{0 < y_{(k-r+1)} < x, y_{(r)} < -x \text{ or } y_{(k-r+1)} > x, y_{(r)} < -y_{(k-r+1)}\right\}\right\} \\
E_4(x) &= \left\{y \in R^{k-1} : y_{(k-r+1)} < 0\right\} \\
E_5(x) &= R^{k-1} - E_1(x) \cup E_{-1}(x) \cup E_a(x) \cup E_{-a}(x)
\end{aligned}$$

where $y_m = Med(x, y_1, \ldots, y_{k-1})$.

To obtain $E_{H_0}(\phi(X_1, X_2, \ldots, X_k)|X_1 = x)$, it is enough to compute the probabilities for the above sets(events) under null by considering the joint distribution of the concerned order statistics from Y_1, Y_2, $\ldots$, Y_{k-1}. The null probabilities are,

$$P\left[E_1(x)\right] = 2^{-(k-1)} \sum_{i=0}^{r-1} \binom{k-1}{i},$$

since $x > 0$, for $E_1(x)$ there can be at most $r - 1$ of Y's negative and under H_0, $P[Y_i > 0] = 1/2$,

$$P\left[E_4(x)\right] = 2^{-(k-1)} \sum_{i=k-r+1}^{k-1} \binom{k-1}{i},$$

since $x > 0$, for $E_4(x)$ there can be at least $k - r - 1$ of Y's negative,

$$P[E_2(x)] = \int_0^x \int_{-x}^0 \int_0^w f(u, v, w)dvdudw + \int_x^\infty \int_{-w}^0 \int_0^w f(u, v, w)dvdudw, \tag{6}$$

where $f(u, v, w)$ is the density functions of $Y_{(r)}$, $Y_m = Med(x, Y_1, \ldots, Y_{k-1})$ and $Y_{(k-r)}$ in a random sample of Y_1, $\ldots$, Y_{k-1} from $G(\cdot)$. Similarly,

$$P[E_3(x)] = \int_0^x \int_{-\infty}^0 \int_{\infty}^{max\{-w,v\}} f(u, v, w)dudvdw + \int_x^\infty \int_{-\infty}^{-w} \int_{-\infty}^{min\{|z|,-x\}} \int_u^{min\{0,z\}} f_1(u, v, z, w)dvdudzdw.$$

where $f(u, v, w)$ is defined in (6) and $f_1(u, v, z, w)$ is the density function of $Y_{(r)}$, $Y_m = Med(x, Y_1, \ldots, Y_{k-1}, Y_{(k-r)})$ and $Y_{(k-r+1)}$ in a random sample of $Y_1, \ldots, Y_{k-1}$ from $G(\cdot)$.

Case II: (When $x < 0$): From Lemma 6.2 we have that $E_1(x) = E_4(-x)$ and $E_2(x) = E_3(-x)$, for $x < 0$, which implies that $P[E_1(x)] = P[E_4(-x)]$ and $P[E_2(x)] = P[E_3(-x)]$. Therefore,

$$E_{H_0}\left(\phi(X_1, \ldots, X_k)|X_1 = x\right) = \begin{cases} P[E_1(x)] - P[E_4(x)] + a\left\{P[E_2(x)] - P[E_3(x)]\right\}, \text{ if } x \geq 0 \\[2ex] P[E_4(-x)] - P[E_1(-x)] + a\left\{P[E_3(-x)] - P[E_2(-x)]\right\}, \text{ if } x \leq 0 \end{cases}$$

Finally,

$$\begin{aligned}
\zeta_1(k, r) &= Var_{H_0}[E_{H_0}\left(\phi(X_1, \ldots, X_k)|X_1 = x\right)] \\
&= \int_0^\infty \left[P[E_1(x)] - P[E_4(x)] + a\left\{P[E_2(x)] - P[E_3(x)]\right\}\right]^2 dG(x) \\
&\quad + \int_{-\infty}^0 \left[P[E_4(-x)] - P[E_1(-x)] + a\left\{P[E_3(-x)] - P[E_2(-x)]\right\}\right]^2 dG(x).
\end{aligned} \tag{7}$$

3.3. Expectation and asymptotic variance for particular cases

One can show that

$$\begin{aligned} & \mu(\theta,5,1) = E[\phi(X_1,X_2,X_3,X_4,X_5,)] \\ = \; & G^5(\theta) - G^5(-\theta) + 5a\left\{(1-2G(-\theta))G(-\theta) + (2-G(-\theta))G^4(-\theta)\right. \\ + & \left.\int_{-\infty}^{-\theta} G(-t-2\theta)\left\{4G^3(t) - 6G^2(t)G(-t-2\theta) + 4G(t)G^2(-t-2\theta) - G^3(-t-2\theta)\right\} dG(t)\right\}, \quad (8) \end{aligned}$$

and

$$Var[V_a(5,1)] = 25\left(\frac{1}{256} + \frac{9a}{320} + \frac{197a^2}{2880}\right).$$

Similarly for $(k,r) = (3,1)$ one can show that,

$$\begin{aligned} \mu(\theta,3,1) &= E[\phi(X_1,X_2,X_3)] \\ &= G^3(\theta) - G^3(-\theta) + 3a\Big\{(1-2G(-\theta))G(-\theta) + G^3(-\theta) \\ &\quad - \int_{-\infty}^{-\theta} G(-t-2\theta)(G(-t-2\theta) - 2G(t))dG(t)\Big\}, \quad (9) \end{aligned}$$

and

$$Var[V_a(3,1)] = 9\left(\frac{1}{16} + \frac{a}{12} + \frac{a^2}{20}\right). \quad (10)$$

4. Performance of the tests based on ARE

Let T be a sequence of test statistics for testing the hypothesis that the median is equal to zero. Let $E(T) = \mu_n(\theta)$ and $Var(T) = \sigma_n^2(\theta)$. Under certain regularity conditions (see Randles and Wolfe (1979), pp. 147-149) the efficacy of T is given by

$$eff[T] = \lim_{n\to\infty} \frac{\mu_n'(0)}{\sqrt{n}\sigma_n(0)}.$$

By considering $T = V_a(k,r)$, $\mu_n(\theta) = \mu(\theta,k,r)$, we will have

$$eff^2\left[V_a(k,r)\right] = \frac{\left[\mu'(0,k,r)\right]^2}{k^2\zeta_1(k,r)}, \quad (11)$$

which depends on the $G(.)$ and the constant a. For given $G(\cdot)$, the optimal value $a^*_{(k,r)}$ of a is obtained by solving $(d/da)eff^2(V_a(k,r)) = 0$ and verifying $(d^2/da^2)eff^2(V_a(k,r)) < 0$ at the solution obtained.

4.1. Efficacies for particular cases

From (9) we have,

$$\mu'(0,3,1) = \frac{3}{2}g(0) + 12aI_1,$$

where $I_1 = \int_{-\infty}^{0}(1-2G(t))g^2(t)dt$, hence from (11),

$$eff^2[V_a(3,1)] = \frac{\left[\mu'(0,3,1)\right]^2}{k^2\zeta_1(3,1)} = \frac{60[g(0)+8aI_1]^2}{(15+20a+12a^2)}$$

and the optimal $a^*_{(3,1)}$ is,

$$a^*_{(3,1)} = \left(\frac{60I_1 - 5g(0)}{6g(0) - 40I_1}\right).$$

Similarly, efficacy of $V_a(5,1)$ is given by,

$$eff^2[V_a(5,1)] = \frac{45\,[2g(0) + a\{5g(0) + 384I_2 - 1\}]^2}{(45 + 324a + 788a^2)},$$

as

$$\mu'(0,5,1) = \frac{5}{8}g(0) + \frac{5}{16}a\{5g(0) + 384I_2 - 1\},$$

where $I_2 = (1/3)\int_{-\infty}^{0}(1 - 2G(t))^3 g^2(t)dt$.
The optimal value $a^*_{(5,1)}$ is,

$$a^*_{(5,1)} = \left(\frac{45 + 99g(0) - 17325I_2}{1576 - 7718g(0) - 606760I_2}\right)$$

In Table 1, optimal values $a^*_{(3,1)}$, $a^*_{(5,1)}$ together with efficacies of $V_{a^*_{(3,1)}}(3,1)$, $V_{a^*_{(5,1)}}(5,1)$ and other competitors for various models are given.

From Table 1, we observe that $V_{a^*_{(3,1)}}(3,1)$ performs better than other competitors for almost all the models. For logistic model efficacy of $V_{a^*_{(3,1)}}(3,1)$ is closer to Wilcoxon signed-rank test, the one known to be locally most powerful for this model.

Under H_0 it is known that $G(\cdot)$ is symmetric about zero. Hence the choice of a should not depend on a specified model. Whatsoever be the symmetric model, we recommend $a = 2.5088$, the optimal value corresponding to the logistic model. For the tests proposed by Mehra et al. (1990) and Shetty and Pandit (2000) we consider the values corresponding to normal model which are respectively, 1.2426 and 2.0933.

From Table 2, we note that still the efficacy of $V_a(3,1)$ with a obtained under the logistic setup continues to be more as compared to sign test and other competitors for various distributions.

In the next section, we perform empirical power study to assess the performances of the proposed tests T_a, $U_a(4,2)$ and $V_a(3,1)$ with t-test, Sign and Wilcoxon signed-rank tests.

5. Performance of the asymptotic tests based on the empirical power

Under H_0, the statistics $V_a(3,1)$ is asymptotically normal with mean 0 and variance given by (10). Thus the criterion to test H_0 versus H_1 at level α is,

$$\text{reject } H_0 \text{ if } \frac{\sqrt{(n)}|V_{a^*_{(3,1)}}(3,1)|}{k\sqrt{\zeta_1(3,1)}} \geq z_{\alpha/2}$$

where $z_{\alpha/2}$ is the upper $(\alpha/2)^{th}$ percentile of standard normal distribution. Similarly, the criterion for the test statistics T_a and $U_a(k,r)$ proposed by Mehra et al. (1990) and Shetty and Pandit (2000) are also defined.

An empirical study was carried out for moderate sample size $n = 25$, using the samples from first seven standard models given in Table 1 and two models from a family of heavy tailed distributions with density $h(x,p)$.

The pdf $h(x,p)$ is defined as

$$h(x,p) = \frac{sin(\frac{\pi}{p})p}{\pi}.\frac{1}{1 + |x|^p} \quad , \ -\infty < x < \infty \text{ and } p > 1.$$

Note that, Cauchy distribution is a member of $h(x, p)$. For the above family of distributions, it appears obtaining optimal value of 'a' that maximizes the efficacy is difficult, as it involves solving the complicated integrals when $p \neq 2$. We propose the test using optimal value of a corresponding to the Cauchy model. In Table 3, empirical powers of the tests for various models are given.

We have also studied the performances of $V_a(3, 1)$ based on empirical power using optimal a corresponding to the logistic model, i.e. $a = 2.5088$. The results are tabulated in the Table 4.

From Table 3, we observe that the empirical power of $V_{a^*_{(3,1)}}(3, 1)$ is higher than the Sign test, which is known to perform better for heavy tailed models. Though for other models, the performance of $V_{a^*_{(3,1)}}(3, 1)$ is not the best, its performance is pretty close to the superior ones from the class of tests considered. From Table 4 we observe that, in practice one can safely use $V_{2.5088}(3, 1)$.

6. Conclusion

In this article, we proposed a class of distribution-free tests for one-sample location problem, which includes test statistics T_a and $U_a(4, 2)$ proposed by Mehra et al. (1990) and Shetty and Pandit (2000) respectively. The proposed test statistics is of U-Statistics type, depends on a constant a and r^{th}, $(k-r+1)^{th}$ order statistics taken from sub-sample of size k together with its median. Expressions for expected value and asymptotic variance are obtained for any (k, r). The optimal value of a is obtained by maximizing the efficacy (in Pitman sense) of the test.

Though the optimal a depends on (k, r), symmetric model $G(\cdot)$, from practical point of view one can safely use $(k, r) = (3, 1)$ and the corresponding optimal value $a^*_{(3,1)} = 2.5088$, obtained under logistic model.

Acknowledgement

Authors deeply acknowledge the Department of Science and Technology, Ministry of HRD, Govt. of India, for the financial support. DST no. SR/S4/MS:373/06.

Appendix

We illustrate the computation of $P[E_i]$ and $P[E_i(x)]$, $i = 1, 2, 3, 4$, only for the case when $(k, r) = (5, 1)$, as the case $(k, r) = (3, 1)$ follows similar steps.

When $k = 5$ and $r = 1$, from (1), (2) we have that $P[E_1] = (1 - G(-\theta))^5$ and $P[E_4] = G^5(-\theta)$. From (3), (4), we have

$$P[E_2] = 5! \int_{-\infty}^{0} \int_{-u}^{\infty} \int_{0}^{w} \{[G(v-\theta) - G(u-\theta)][G(w-\theta) - G(v-\theta)]\} \, dG(v-\theta) dG(w-\theta) dG(u-\theta).$$

Thus integrating w.r.t. v and w, we get,

$$\begin{aligned} P[E_2] &= 10G^4(-\theta) - 10G^2(-\theta) + 5G(-\theta) - \int_{-\infty}^{-\theta} \left[5G^4(-t-2\theta) - 20G(t)G^3(-t-\theta)\right. \\ &\quad \left. - \left(G^2(-\theta) - 2G(t)G(-\theta)\right) G^2(-t-\theta) + 20\left(2G^3(-\theta) - 3G(t)G^2(-\theta)\right) G(-t-\theta)\right] dG(t) \end{aligned}$$

and

$$\begin{aligned} P[E_3] &= 5G^5(-\theta) + \int_{-\infty}^{-\theta} \left[30\left(G^2(-\theta) - 2G(t)G(-\theta) + G^2(t)\right) G^2(-t-\theta)\right. \\ &\quad \left. -20\left(2G^3(-\theta) - 3G(t)G^2(-\theta) + G^3(t)\right) G(-t-\theta)\right] dG(t) \end{aligned}$$

Now, substituting the above probabilities in (5) yields (8).

Next, we compute $P[E_i(x)]$, $i = 1, 2, 3, 4$ under null hypothesis. We have $P[E_1(x) = 1/16$ and $P[E_4(x)] = 0$. To obtain $P[E_2(x)]$, $P[E_3(x)]$, substituting the joint density function of Y_1, Y_3 and Y_4 obtained from a random sample of Y_1, Y_2, Y_3,Y_4, in (6), we get,

$$\begin{aligned}
P[E_2(x)] &= 4!\left\{\int_0^x\int_{-x}^0\int_0^w [G(v)-G(u)]\,dG(v)dG(u)dG(w) + \int_x^\infty\int_{-w}^0\int_0^w [G(v)-G(u)]\,dG(v)dG(u)dG(w)\right\}\\
&= 4G^4(x) - 8G^3(x) + 6G^2(x) - 2G(x) + (5/8).
\end{aligned}$$

and

$$\begin{aligned}
P[E_3(x)] &= 4!\left\{\int_x^\infty\int_{-\infty}^{-w}\int_u^0 [G(v)-G(u)]\,dG(v)dG(u)dG(w) + \int_0^x\int_{-\infty}^{-x}\int_u^0 [G(v)-G(u)]\,dG(v)dG(u)dG(w)\right.\\
&\quad\left. + \int_{-x}^0\int_u^w\int_u^0 [G(v)-G(u)]\,dG(v)dG(u)dG(w) + \int_{-\infty}^{-x}\int_u^w\int_u^0 [G(v)-G(u)]\,dG(v)dG(u)dG(w)\right\}\\
&= -4G^4(x) + 8G^3(x) - 6G^2(x) + 2G(x).
\end{aligned}$$

Similarly, for $x < 0$, from Lemma 6.2, if we replace x by $-x$ in the above we get $P[E_1(-x) = 0$ and $P[E_4(-x)] = 1/16$,

$$P[E_2(-x)] = -4G^4(x) + 8G^3(x) - 6G^2(x) + 2G(x),$$

$$P[E_3(-x)] = 4G^4(x) - 8G^3(x) + 6G^2(x) - 2G(x) + (5/8).$$

Thus, putting $P[E_i(x)]$ and $P[E_i(-x)]$, $i = 1, 2, 3, 4$, in (7), we get,

$$\begin{aligned}
\zeta_1(5,1) &= \int_0^\infty \left[(1/16) + a\left\{8G^4(x) - 16G^3(x) + 12G^2(x) - 4G(x) + (5/8)\right\}\right]^2 dG(x)\\
&\quad + \int_{-\infty}^0 \left[a\left\{-8G^4(x) + 16G^3(x) - 12G^2(x) + 4G(x) - (5/8)\right\} - (1/16)\right]^2 dG(x),\\
&= \frac{1}{256} + \frac{9a}{320} + \frac{197a^2}{2880}.
\end{aligned}$$

Lemma 6.1. *Following are the relations between probabilities of the events E_i, $i = 1, 2, 3, 4$,*

(i) $P_\theta[E_1] = P_{-\theta}[E_4]$,

(ii) $P_\theta[E_2] = P_{-\theta}[E_3]$.

Proof. (i) From (1), we have,

$$P_\theta[E_1] = 1 - \int_0^{G(-\theta)} \frac{x^{r-1}(1-x)^{(k-r)}}{\beta(r, k-r+1)}dx = \int_{G(-\theta)}^1 \frac{x^{r-1}(1-x)^{(k-r)}}{\beta(r, k-r+1)}dx.$$

Putting $1 - x = y$, we have

$$P_\theta[E_1] = \int_0^{G(\theta)} \frac{y^{k-r}(1-y)^{(r-1)}}{\beta(k-r+1, r)}dy = P_{-\theta}[E_4].$$

(ii) This can be established similarly by transforming u, v, w to $-x$, $-y$ and $-z$ respectively in (4).

Hence the proof. □

Lemma 6.2. *Under H_0, when $x > 0$ and for the events $E_i(x)$, $i = 1, 2, 3, 4$, we have*

(i) $P_{H_0}[E_2(x)] = P_{H_0}[E_3(-x)]$,

(ii) $P_{H_0}[E_1(x)] = P_{H_0}[E_4(-x)]$.

Proof. (i) For simplicity we give the proof when $x < 0$ and $k = 2m - 1$, as a consequence of this, one can prove for $x > 0$.

We have,

$$E_2(x) = \Big\{\{y_{(m-1)} > 0\} : y_{(r-1)} < 0, y_{(k-r)} > 0 : \{y_{(r)} < x\} \cap \{y_{(k-r)} > y_{(r)}\}$$
$$\text{or } \{x < y_{(r)} < 0\} \cap \{y_{(k-r)} > -\max\{y_{(r-1)}, x\}\Big\}.$$

By using the fact that, Y and $-Y$ have same distribution, transforming $-Y_r$ to $U_{(k-r)}$, putting $t = -x$, we have

$$E_2(-t) = \Big\{\{u_{(m)} < 0\} : u_{(r)} < 0, u_{(k-r+1)} > 0 : \{u_{(k-r)} > t\} \cap \{u_{(r)} < u_{(k-r)}\}$$
$$\text{or } \{t > u_{(k-r)} < 0\} \cap \{u_{(r)} < max\{u_{(k-r+1)}, t\}\Big\} = E_3(t).$$

This implies $P[E_2(x)] = P[E_3(-x)]$, when $x < 0$.

(ii) One can establish this similar to the above.

Hence the proof.
□

References

Bandyopadhyay, U., Datta, D. (2007) Adaptive Nonparametric tests for Single sample location problem, Statistical Methodology 4, 423–433.

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

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

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

Mehra, K. L., Prasad, N. G. N., Madhava Rao, K. S. (1990) A Class of Nonparametric Tests for the One-Sample Location Problem, Australian Journal of Statistics 32, 373–392.

Pandit, P. V., Math, C. M. (2011) A Nonparametric Control Chart for Location based on Sub-Samples of size two, Economic Quality Control 26, 91–98.

Randles, R. H., Wolfe, D. A. (1979) Introduction to the Theory of Nonparametric Statistics, John Wiley & Sons, New York.

Shetty, I. D., Pandit, P. V. (2000) A class of Distribution-free Tests for One-Sample Location Problem, Journal of the Karnataka University-Science 43, 36–41.

Certain corrections in this paper has been uploaded as the file, Corrections to PSF-2012-13.

Table 1: Efficacies of t-test, Sign test (B), Wilcoxon Signed-rank test (W), T_{a^*}, $U_{a^*}(4,2)$, $V_{a^*}(3,1)$ and $V_{a^*}(5,1)$ for few Standard Distributions

Models	$a^*_{(3,1)}$	$a^*_{(5,1)}$	$eff^2(t)$	$eff^2(B)$	$eff^2(W)$	$eff^2(T_{a^*})$	$eff^2(U_{a^*}(4,2))$	$eff^2(V_{a^*_{(3,1)}}(3,1))$	$eff^2(V_{a^*_{(5,1)}}(5,1))$
Cauchy	-0.1762	-0.0230	-	0.4053	0.3040	0.4053	0.4252	0.6411	0.4369
Laplace	0	0	1	2.0000	1.5000	2.0000	2.0000	2.0000	2.0000
Logistic	2.5088	0.0033	1	0.8225	1.0966	1.0966	1.1364	1.0799	0.8247
Normal	7.2960	0.0051	1	0.6366	0.9549	0.9643	0.9503	0.9868	0.6414
Parabolic	-2.4991	-0.0149	1	0.4500	0.8460	0.9360	0.8125	1.0829	1.5956
Triangular	4.4766	0.0059	1	0.6667	0.8889	0.8889	0.7965	0.9657	0.6735
Uniform	-2.4991	0.0180	1	0.3333	1.3333	1.3333	0.6869	2.0021	0.4041
$\frac{1}{2}\lvert y\rvert I_{\lvert y\rvert\leq\sqrt{2}}$	-1.5	0.0286	1	0.0000	2.6667	10.6600	16.4563	14.3957	0.1031

Table 2: Efficacies of ($T_{1.2426}$), ($U_{2.0938}(4,2)$) and $V_{2.5088}(3,1)$ for few Standard modles

Models	$eff^2(T_{(1.2426)})$	$eff^2(U_{(2.0938)}(4,2))$	$eff^2(V_{(2.5088)}(3,1))$
Cauchy	0.3676	0.1323	0.2688
Laplace	1.3204	0.1011	1.5227
Logistic	1.0859	0.3572	1.0799
Normal	0.9643	0.9503	0.9707
Parabolic	0.9052	0.3572	0.9486
Triangular	0.8803	0.1916	0.9595
Uniform	1.1068	0.1771	1.2878
$\frac{1}{2}\lvert y\rvert I_{\lvert y\rvert\leq\sqrt{2}}$	3.6247	0.1771	3.4344

Table 3: Empirical Power of the Tests for various models with $\alpha = 0.05$, $n = 25$ and number of Monte Carlo simulations is 10000.

Models	Tests	θ= 0	θ=0.2	θ=0.4	θ=0.6	θ=0.8	θ= 1.0
Cauchy	t	0.019	0.026	0.052	0.091	0.145	0.202
	B	0.042	0.083	0.206	0.386	0.583	0.735
	W	0.046	0.083	0.175	0.321	0.476	0.618
	T_{a^*}	0.042	0.083	0.206	0.386	0.584	0.735
	$U_{a^*_{(4,2)}}(4,2)$	0.046	0.092	0.220	0.406	0.604	0.754
	$V_{a^*_{(3,1)}}(3,1)$	0.049	0.095	0.228	0.412	0.606	0.755
Laplace	t	0.045	0.077	0.171	0.325	0.513	0.683
	B	0.044	0.094	0.211	0.389	0.579	0.736
	W	0.045	0.091	0.212	0.392	0.594	0.752
	T_{a^*}	0.044	0.094	0.211	0.389	0.578	0.736
	$U_{a^*_{(4,2)}}(4,2)$	0.044	0.094	0.211	0.389	0.578	0.736
	$V_{a^*_{(3,1)}}(3,1)$	0.044	0.094	0.211	0.389	0.578	0.736
Logistic	t	0.046	0.155	0.499	0.825	0.965	0.996
	B	0.042	0.125	0.409	0.722	0.914	0.982
	W	0.046	0.160	0.517	0.840	0.971	0.997
	T_{a^*}	0.050	0.169	0.533	0.848	0.972	0.998
	$U_{a^*_{(4,2)}}(4,2)$	0.062	0.144	0.523	0.829	0.967	0.996
	$V_{a^*_{(3,1)}}(3,1)$	0.046	0.154	0.500	0.822	0.959	0.993
Normal	t	0.053	0.165	0.488	0.814	0.969	0.998
	B	0.048	0.113	0.329	0.623	0.859	0.965
	W	0.052	0.151	0.467	0.796	0.963	0.997
	T_{a^*}	0.054	0.161	0.478	0.807	0.963	0.997
	$U_{a^*_{(4,2)}}(4,2)$	0.058	0.170	0.462	0.765	0.942	0.992
	$V_{a^*_{(3,1)}}(3,1)$	0.053	0.153	0.445	0.762	0.901	0.980
Parabolic	t	0.051	0.499	0.984	1.000	1.000	1.000
	B	0.042	0.167	0.628	0.981	1.000	1.000
	W	0.046	0.451	0.944	1.000	1.000	1.000
	T_{a^*}	0.051	0.590	0.976	1.000	1.000	1.000
	$U_{a^*_{(4,2)}}(4,2)$	0.055	0.398	0.898	1.000	1.000	1.000
	$V_{a^*_{(3,1)}}(3,1)$	0.049	0.622	0.977	1.000	1.000	1.000
Triangular	t	0.051	0.156	0.477	0.826	0.973	0.999
	B	0.044	0.110	0.301	0.577	0.816	0.946
	W	0.049	0.145	0.436	0.780	0.956	0.997
	T_{a^*}	0.053	0.154	0.453	0.796	0.962	0.997
	$U_{a^*_{(4,2)}}(4,2)$	0.058	0.143	0.405	0.732	0.931	0.991
	$V_{a^*_{(3,1)}}(3,1)$	0.048	0.149	0.457	0.810	0.961	0.998

Models	Tests	θ= 0	θ=0.2	θ=0.4	θ=0.6	θ=0.8	θ= 1.0
Uniform	t	0.050	0.154	0.470	0.823	0.976	0.998
	B	0.042	0.080	0.191	0.392	0.659	0.863
	W	0.046	0.143	0.417	0.743	0.939	0.992
	T_{a^*}	0.053	0.183	0.500	0.816	0.962	0.996
	$U_{a^*_{(4,2)}}(4,2)$	0.051	0.154	0.357	0.644	0.863	0.976
	$V_{a^*_{(3,1)}}(3,1)$	0.046	0.173	0.497	0.816	0.961	0.997
$h(x,1.4)$	t	0.048	0.093	0.223	0.445	0.690	0.866
	B	0.042	0.092	0.220	0.404	0.605	0.748
	W	0.046	0.096	0.228	0.441	0.672	0.842
	T_{a^*}	0.042	0.092	0.220	0.404	0.605	0.748
	$U_{a^*_{(4,2)}}(4,2)$	0.046	0.099	0.238	0.451	0.669	0.824
	$V_a(3,1)$	0.049	0.100	0.230	0.410	0.607	0.758
$h(x,2.2)$	t	0.049	0.546	0.992	1.000	1.000	1.000
	B	0.045	0.260	0.796	0.994	1.000	1.000
	W	0.047	0.481	0.974	1.000	1.000	1.000
	T_{a^*}	0.045	0.260	0.796	0.994	1.000	1.000
	$U_{a^*_{(4,2)}}(4,2)$	0.049	0.342	0.898	0.995	1.000	1.000
	$V_a(3,1)$	0.051	0.261	0.784	0.992	1.000	1.000

Table 4: Empirical power of ($T_{1.22426}$), ($U_{2.0938}(4,2)$) and $V_{2.5088}(3,1)$ with $\alpha = 0.05$, $n = 25$ and number of Monte Carlo simulations is 10000.

Models	Tests	θ= 0	θ=0.2	θ=0.4	θ=0.6	θ=0.8	θ= 1.0
Cauchy	$T_{1.22426}$	0.050	0.082	0.162	0.292	0.431	0.553
	$U_{2.0938}(4,2)$	0.054	0.105	0.209	0.367	0.530	0.654
	$V_{2.5088}(3,1)$	0.046	0.074	0.141	0.245	0.353	0.449
Laplace	$T_{1.22426}$	0.056	0.088	0.214	0.373	0.571	0.733
	$U_{2.0938}(4,2)$	0.058	0.082	0.183	0.322	0.477	0.596
	$V_{2.5088}(3,1)$	0.051	0.084	0.200	0.342	0.528	0.684
Parabolic	$T_{1.22426}$	0.052	0.590	0.976	1.000	1.000	1.000
	$U_{2.0938}(4,2)$	0.058	0.398	0.898	1.000	1.000	1.000
	$V_{2.5088}(3,1)$	0.046	0.548	0.967	1.000	1.000	1.000
Triangular	$T_{1.22426}$	0.048	0.149	0.448	0.798	0.961	0.998
	$U_{2.0938}(4,2)$	0.053	0.118	0.322	0.575	0.733	0.784
	$V_{2.5088}(3,1)$	0.046	0.149	0.451	0.808	0.966	0.998
Uniform	$T_{1.22426}$	0.049	0.165	0.474	0.798	0.958	0.996
	$U_{2.0938}(4,2)$	0.053	0.139	0.353	0.656	0.887	0.976
	$V_{2.5088}(3,1)$	0.046	0.174	0.497	0.817	0.961	0.997
$h(x,1.4)$	$T_{1.22426}$	0.051	0.080	0.218	0.438	0.606	0.764
	$U_{2.0938}(4,2)$	0.050	0.105	0.238	0.442	0.673	0.824
	$V_{2.5088}(3,1)$	0.050	0.091	0.248	0.455	0.687	0.850
$h(x,2.2)$	$T_{1.22426}$	0.052	0.524	0.982	1.000	1.000	1.000
	$U_{2.0938}(4,2)$	0.053	0.342	0.764	0.803	0.954	1.000
	$V_{2.5088}(3,1)$	0.049	0.370	0.899	0.998	1.000	1.000