Journal of Statistics Applications & Probability
--- *An International Journal*

A Note on Test of Homogeneity Against Umbrella Scale Alternative Based on *U*-Statistics

Anil Gaur

Department of Statistics, Panjab University, Chandigarh, India
Email Address: anilgaur8@gmail.com

Received July7, 2012; Revised September 28, 2012; Accepted October 2, 2012

Abstract: A fundamental problem encountered in statistics is that of testing the equality of scale parameters against umbrella alternative with at least one strict inequality. In this article, a nonparametric test based on *U*-statistic by considering the subsample minima and maxima for several sample scale problem against umbrella alternative, when peak of the umbrella is known, is proposed. The proposed test have the advantage of not requiring the several distribution functions to have a common median, but rather any common quantile of order q, $0 \le q \le 1$, (not necessarily ½) which is assumed to be known. Pitman efficiency indicate that the proposed test is equivalent to the test B proposed by Gaur, Mahajan and Arora (2012).

Keywords: Umbrella ordering, *U*-statistic, Test for scale, Asymptotic Relative Efficiency.

1 Introduction

Let $X_{i1}, X_{i2}, ..., X_{in_i}; i = 1, 2, ..., k$, be independent random samples of size n_i from absolutely continuous cumulative distribution functions $F_i(x) = F\left(\frac{x}{\theta_i}\right)$, $i = 1, 2, ..., k$. We assume that these distribution functions have θ_0 as the common quantile of order $q, (0 \le q \le 1)$ *i.e.*, $F_i(\theta_0) = q$ for $i = 1, 2, ..., k$. Without loss of generality we assume that the common known quantile $\theta_0 = F_1^{-1}(q) = ... = F_k^{-1}(q)$ of order q is zero for the pre-specified q. It is also assumed that $F_i(x)$, $i = 1, 2, ..., k$, are identical in all respects except possibly their scale parameters.

The problem of testing the null hypothesis of homogeneity of scale parameters $H_0 : \theta_1 = ... = \theta_k$ against simple ordered alternative hypothesis $H_A : \theta_1 \le ...... \le \theta_k$ with at least one strict inequality has received considerable attention in the literature. For some earlier work on this problem, one may refer to Rao [16], Shanubhogue [18], Kusum and Bagai [11], Gill and Dhawan [9], Singh and Gill [21] among others. Most of the tests, except Kusum and Bagai [11] test, require the assumption that the common quantile of different distributions is of order q = ½, *i.e.*, the distributions have the same median. None of these tests is adequate when the common quantile is different from median. The asymmetry of the situation is not reflected in the statistics used in the above tests. Kusum and Bagai [11] considered a more general version of this problem. In the traditional set up, the dispersion of a distribution is evaluated around its central value (say the median). However, in many applications, the dispersion needs to be considered around a non-central value, say a quantile of order q, $0 \le q \le 1$, (not necessarily ½) which is assumed to be known. These procedures have important applications for problems where the treatments can be assumed to satisfy a simple ordering, such as for a sequence of increasing dose-level of a drug.

Umbrella ordering is important in dose-response experiment (e.g., see Simpson and Margolin [20]).

In case where mode of action of a drug is related to its toxic effects, e.g., in case of life saving therapy of heart failure, life saving digitalization therapy of heart failure, umbrella behavior is anticipated and careful dosage planning is required.

There has been substantial work in testing equality of location parameters against umbrella alternative with at least one strict inequality, but very little work for testing of scale parameters. For detailed references one may refer, Mack and Wolfe [13], Chen and Wolfe [5], Chen and Wolfe [6], Chen [4], Kosseler [10] and Abebe and Singh [1]. Shetty and Bhat [19], Bhat and Patil [3] and Bhat [2] proposed test statistics based on linear combination of two sample *U*-statistics for testing homogeneity of location parameters against umbrella alternative with at least one strict inequality.

Singh and Liu [22] proposed a test statistics for homogeneity of scale parameters against umbrella alternative with at least one strict inequality based isotonic estimator of scale parameter. They also provided one-sided simultaneous confidence intervals for all the ordered pairwise scale ratios, and critical points for two parameter exponential probability distribution. Recently, Gaur et al. [8] provided three test statistics based on weighted linear combination of two-sample *U*-statistics for testing homogeneity of scale parameters against umbrella alternative, with at least one strict inequality, when the peak of the umbrella, *h* is known.

In this paper, a new test based on subsample minima and maxima for testing homogeneity of scale parameters against umbrella alternative, with at least one strict inequality, is proposed when peak of the umbrella, *h* is known. It is found that the proposed test is equivalent to the test *B* proposed by Gaur et al. [8] for heavy-tailed distributions, when the common quantile is not the median but of some order *q* (*q* not necessarily equal to ½). Situations where two populations may have a common quantile of order other than *q* = ½ arise in many real life examples. In particular, this assumption appears to be quite realistic, as in examples of automatic can filling mechanisms and models of wage distribution as pointed by Deshpande and Kusum [7].

This article is organized as follows. The new proposed test for testing homogeneity of scale parameters against umbrella alternative, with at least one strict inequality and its distribution is given in section 2. Section 3 is devoted to the optimal choice of weights. Efficacy results and Pitman asymptotic relative efficiency are given in section 4.

2 The Proposed Test and its Distribution

Let $X_{i1}, X_{i2}, \ldots\ldots., X_{in_i}; i = 1, 2, \ldots\ldots., k$, be k independent random samples of size n_i from absolutely continuous cumulative distribution functions $F_i(x) = F\left(\frac{x}{\theta_i}\right), i = 1, 2, \ldots\ldots, k$. We assume that these distribution functions have zero as the common quantile of order q, $(0 \le q \le 1)$, *i.e.*, $F_i(0) = q$ for $i = 1, 2, \ldots\ldots, k$. It is also assumed that $\mathrm{F}_i(x)$, $i = 1, 2, \ldots\ldots, k$, are identical in all respects except possibly their scale parameters. The hypothesis, which is of interest in this paper, could be formally stated as follows:

$$H_0 : \theta_1 = \theta_2 = \ldots\ldots\ldots = \theta_k$$

against the umbrella alternative

$$H_1 : \theta_1 \le \theta_2 \le \ldots. \le \theta_{h-1} \le \theta_h \ge \theta_{h+1} \ge \ldots. \ge \theta_{k-1} \ge \theta_k ,$$

with at least one strict inequality and *h*, the peak of the umbrella, is known.

First we propose a two-sample *U*-statistic in the context of a two-sample scale problem where the assumption of the common quantile of order $q, (0 \le q \le 1)$ is made and then extend it to the *k*-sample problem considered here. Now, for $i < j$, define the kernel

$$\phi_{ij}(X_{i1}, X_{i2}, X_{i3}; X_{j1}, X_{j2}, X_{j3})$$

$$= \begin{cases} 1, & \text{if } 0 \le \max(X_{i1}, X_{i2}, X_{i3}) \le \min(X_{j1}, X_{j2}, X_{j3}) \text{ and } X_{i1}, X_{i2}, X_{i3}, X_{j1}, X_{j2}, X_{j3} \ge 0 \\ & or\ \max(X_{j1}, X_{j2}, X_{j3}) \le \min(X_{i1}, X_{i2}, X_{i3}) < 0 \text{ and } X_{i1}, X_{i2}, X_{i3}, X_{j1}, X_{j2}, X_{j3} \le 0, \\ -1, & \text{if } 0 \le \max(X_{j1}, X_{j2}, X_{j3}) \le \min(X_{i1}, X_{i2}, X_{i3}) \text{ and } X_{i1}, X_{i2}, X_{i3}, X_{j1}, X_{j2}, X_{j3} \ge 0 \\ & or\ \max(X_{i1}, X_{i2}, X_{i3}) \le \min(X_{j1}, X_{j2}, X_{j3}) < 0 \text{ and } X_{i1}, X_{i2}, X_{i3}, X_{j1}, X_{j2}, X_{j3} \le 0, \\ 0, & otherwise. \end{cases}$$

The two-sample *U*-statistic corresponding to the kernel ϕ_{ij} is

$$U_{ij}(X_{i1}, X_{i2}, X_{i3}; X_{j1}, X_{j2}, X_{j3}) = \frac{1}{\binom{n_i}{3}\binom{n_j}{3}} \sum_c \phi_{ij}(X_{i1}, X_{i2}, X_{i3}; X_{j1}, X_{j2}, X_{j3}) \tag{2.1}$$

where *c* denotes the summation extended over all possible $\binom{n_i}{3}\binom{n_j}{3}$ combinations of $X_{i1}, X_{i2}, \ldots\ldots, X_{in_i}$ and $X_{j1}, X_{j2}, \ldots\ldots, X_{jn_j}$.

The statistic U_{ij} is obviously a *U*-statistic (Lehman [12]) corresponding to the kernel ϕ_{ij}. It can be seen that the kernel takes on a non-zero value only when both the X_i*'s* and X_j*'s* have the same sign. Under H_1, the observations from the i^{th} population are expected to be smaller (larger) than those from the $(i+1)^{th}$ population, therefore $U_{i,i+1}$ ($U_{i+1,i}$), $i = 1, 2, \ldots\ldots, h-1$ ($i = h, h+1, \ldots\ldots, k-1$) is expected to take large values. Motivated by the fact that under H_1, $U_{i,i+1}$ ($U_{i+1,i}$), $i = 1, 2, \ldots\ldots, h-1$ ($i = h, h+1, \ldots\ldots, k-1$) is expected to take large values, we propose a class of test statistics U_k based on subsample extreme of size three as

$$T = T_1 + T_2 \tag{2.2}$$

where, $T_1 = \sum_{i=1}^{h-1} a_i U_{i,i+1}$, $T_2 = \sum_{i=h}^{k-1} a_i U_{i+1,i}$,

for testing H_0 against H_1, where $(a_1, a_2, \ldots\ldots\ldots, a_{k-1})$ are suitable chosen real positive constants. It may be noted that for each set of values $(a_1, a_2, \ldots\ldots\ldots, a_{k-1})$, we get a distinct member of this class of test statistic. A large value of *T* leads to the rejection of H_0 against H_1. When a_i = 1, ($i = 1, 2, \ldots\ldots, k-1$), we obtain Mack-Wolfe version of *T* as

$$T_M = \sum_{i=1}^{h-1} U_{i,i+1} + \sum_{i=h}^{k-1} U_{i+1,i} . \tag{2.3}$$

2.1 The Distribution of T

Clearly,

$$E(T)=\sum_{i=1}^{h-1} a_i \mu_{i,i+1}+\sum_{i=h}^{k-1} a_i \mu_{i+1,i}\ ,$$

where $\mu_{i,\,i+1}=\pi_{i1}-\pi_{i2}$

$$\mu_{i+1,i}=\pi_{i+1,1}-\pi_{i+1,2}\ \text{, and}$$

$$\pi_{i1}=3\int_0^{\infty}[F_i(x)-q]^3[1-F_{i+1}(x)]^2\,dF_{i+1}(x)+3\int_{-\infty}^{0}[q-F_i(x)]^3F_{i+1}^2(x)dF_{i+1}(x)\,,$$

$$\pi_{i2}=3\int_0^{\infty}[F_{i+1}(x)-q]^3[1-F_i(x)]^2\,dF_i(x)+3\int_{-\infty}^{0}[q-F_{i+1}(x)]^3F_i^2(x)dF_i(x)\,,$$

$$\pi_{i+1,1}=3\int_0^{\infty}[F_{i+1}(x)-q]^3[1-F_i(x)]^2\,dF_{i+1}(x)+3\int_{-\infty}^{0}[q-F_{i+1}(x)]^3F_i^2(x)dF_i(x)\,,$$

$$\pi_{i+1,2}=3\int_0^{\infty}[F_i(x)-q]^3[1-F_{i+1}(x)]^2\,dF_{i+1}(x)+3\int_{-\infty}^{0}[q-F_i(x)]^3F_{i+1}^2(x)dF_{i+1}(x)\,.$$

Under H_0, we have $F_i(.)=F_j(.)$ which implies $\mu_{i,\,i+1}(\mu_{\,i+1,i})=0,\quad \forall\, i=1,2,\ldots\ldots,h-1$ $(i=h,h+1,\ldots\ldots,k-1)$.

Hence, under H_0, $E(T)=E(T_M)=0$. (2.4)

Using the results of Lehman [12] and Puri [14], the proof of the following theorem follows from the transformation theorem (see Serfling [17] , page 122) immediately.

Theorem 2.1 Let $N=\sum_{i=1}^{k} n_i$. The asymptotic null distribution of $\sqrt{N}\,T$, as $N\to\infty$ in such a way that $\frac{n_i}{N}\to p_i,\ 0<p_i<1,\ i=1,2,\ldots,k,$ is normal with mean 0 and variance η , where

$$\eta=Var(A)=\eta_1+\eta_2+2\eta_{12}\,, \tag{2.5}$$

$$\eta_1=Var(A_1)=\underline{a_1}'\Sigma_1\,\underline{a_1}\ ,\ \ \eta_2=Var(A_2)=\underline{a_2}'\Sigma_2\,\underline{a_2}\ , \tag{2.6}$$

$$\Sigma_1=((\sigma_{1,ij})),\ \Sigma_2=((\sigma_{2,ij}))\,,$$

$$\underline{a_1}'=(a_1,a_2,\ldots\ldots,a_{h-1}),\ \ \underline{a_2}'=(a_h,a_{h+1},\ldots\ldots,a_{k-1})\,,$$

$$\sigma_{1,ij}=\begin{cases}\left(\dfrac{1}{p_i}+\dfrac{1}{p_{i+1}}\right)\xi & \text{for } i=j=1,2,\ldots,h\text{-}1,\\[2ex] -\dfrac{1}{p_{i+1}}\xi & \text{for } j=i+1;\ i=1,2,\ldots\ldots,h\text{-}2,\\[2ex] -\dfrac{1}{p_i}\xi & \text{for } j=i\text{-}1;\ i=2,3,\ldots\ldots,h\text{-}1,\\[1ex] 0 & \text{otherwise,}\end{cases} \tag{2.7}$$

$$\sigma_{2,ij}=\begin{cases}\left(\dfrac{1}{p_i}+\dfrac{1}{p_{i+1}}\right)\xi & \text{for } i=j=h,h+1,\ldots,k\text{-}1,\\ -\dfrac{1}{p_{i+1}}\xi & \text{for } j=i+1;\ i=h,h+1,\ldots\ldots,k\text{-}2,\\ -\dfrac{1}{p_i}\xi & \text{for } j=i\text{-}1;\ i=h+1,h+2,\ldots\ldots,k\text{-}1,\\ 0 & \text{otherwise,}\end{cases} \tag{2.8}$$

$$\eta_{12}=Cov(A_1,A_2)=a_{h-1}a_hCov(T_{h-1,h},T_{h+1,h})=((a_{h-1}a_h)/p_h)\xi\,, \tag{2.9}$$

where $\xi=\dfrac{131}{23100}[q^{11}+(1-q)^{11}]$.

In case all the sample sizes are equal *i.e.*, $p_1=p_2=\ldots\ldots=p_k=\dfrac{1}{k}$, then substituting (2.6), (2.7), (2.8) and (2.9) in (2.5), we have

$$\begin{aligned}\eta&=2k\xi\left[\sum_{i=1}^{h-1}a_i^2-\sum_{i=1}^{h-2}a_ia_{i+1}+\sum_{i=h}^{k-1}a_i^2-\sum_{i=h}^{k-2}a_ia_{i+1}+a_{h-1}a_h\right]\\&=2k\xi\left[\sum_{i=1}^{k-1}a_i^2-\sum_{i=1}^{k-2}a_ia_{i+1}+2a_{h-1}a_h\right].\end{aligned} \tag{2.10}$$

Similarly, the asymptotic null distribution of $\sqrt{N}\,[T_M-E(T_M)]$ is normal with mean zero and variance $\eta_M=6k\xi$.

Practical implementation of this procedure may require an estimator for the common quantile θ_0, under the null hypothesis, of order $\alpha=F_1(\theta_0)=\ldots\ldots=F_k(\theta_0)$. We suggest to use a pooled estimator of $\theta_0=F_1^{-1}(q)=\ldots..=F_k^{-1}(q)$ for a given (predetermined values of) q. To achieve this, we pool all the observations $X_{i1},X_{i2},\ldots\ldots.,X_{in_i};i=1,2,\ldots\ldots.,k$ into a single vector $\boldsymbol{Z}$ and estimate θ_0 by obtaining the *q-th* quantile of $\boldsymbol{Z}$.

Remark 2.1 The corresponding two-sample problem is under consideration in different paper by the author.

3 Optimal Choice of Weights

Under the sequence of Pitman alternatives, the square of the efficacy of test T is given by

$$e(T)=\frac{I^2\left(\sum_{i=1}^{k}a_i\right)^2}{\eta} \tag{3.1}$$

where, $I=18\left\{\int_{-\infty}^{0}x\,F^2(x)[q-F(x)]^2f^2(x)\,dx-\int_{0}^{\infty}x[F(x)-q]^2[1-F(x)]^2f^2(x)\,dx\right\}$.

For efficiency comparisons, we consider the equal sample size and equally spaced alternatives of the type

$$\theta_i = \begin{cases} i\theta & \text{for } i = 1, 2, \ldots, h, \\ (2h-i)\theta & \text{for } i = h+1, \ldots, k \end{cases}, \quad \theta > 0. \tag{3.2}$$

Making use of the results due to Rao [15] (page 60) for determining optimal weights, we obtain the optimal weights a_i^* for which T has maximum efficiency. For odd k and $h=(k+1)/2$,

$$a_i^* = \begin{cases} (i/k)(k(h\text{-}i)\text{-}(h\text{-}1)) & for\ i = 1, 2, \ldots, h\text{-}1, \\ ((k-i)/k)\ (k(i+1\text{-}h)\text{-}(h\text{-}1)) & for\ i = h, \ldots, k\text{-}1. \end{cases} \tag{3.3}$$

The respective square of the efficacy of tests T with optimal choice of weights in (3.3) is given by

$$e^*(T) = \frac{I\left[(k^4+2k^2-3)/4k^2\right]}{12\xi}. \tag{3.4}$$

And the square of the efficacy of tests T_M is given by

$$e^*(T_M) = \frac{I(k-1)^2}{6k\xi}. \tag{3.5}$$

4 Asymptotic Relative Efficiency

We compute the Pitman asymptotic relative efficiency (ARE) of the proposed tests with Gaur-Mahajan-Arora B-test (see Gaur et al. [8]). The efficiency of the new proposed test is given in (3.4) and (3.5). The efficacy of B is given by

$$e^*(B) = \frac{J\left[(k^4+2k^2-3)/4k^2\right]}{12\zeta},$$

$$e^*(B_M) = \frac{J(k-1)^2}{6k\zeta},$$

where $\zeta = \frac{1572}{1925}\left[q^{11} + (1-q)^{11}\right]$ and

$$J = 72\left\{\int_{-\infty}^{0} x\,F^2(x)[q-F(x)]^2 f^2(x)\,dx - \int_{0}^{\infty} x\,[F(x)-q]^2[1-F(x)]^2 f^2(x)\,dx\right\}.$$

Then the asymptotic relative efficiency (ARE) of T (T_M) with respect to B (B_M) test can be computed from the ratio of the Pitman efficacies, and we notice that

$$\text{ARE}(T, B) = \text{ARE}(T_M, B_M) = 1.$$

Also, $ARE\ (T, T_M) = \dfrac{(k^4+2k^2-3)}{8k(k-1)^2}$ and the values of ARE for different values of k can be found in Gaur et al [8].

Gaur et. al. [8] mentioned that the test B which is based on the medians of subsamples performs better for heavy-tailed distribution (when the different distributions have common quantile). Since asymptotic relative efficiency of the test T with respect to test B is 1, therefore the test T which is based on the minima and maxima of subsamples performs better for heavy-tailed distribution and is equivalent to B-test. So, in case of heavy-tailed distributions like the Double Exponential distribution and Cauchy distribution when the distributions have common quantile then the test T can be used in place of B for testing the homogeneity of scale parameters against umbrella alternative with at least one strict inequality.

Remark 4.1 It can be noted that the proposed test has less variance then B-test of Gaur et al. [8]. Also, the asymptotic variance computation of the proposed test is simpler than the asymptotic variance computation of B-test of Gaur et al. [8].

Acknowledgments

The author is thankful to the anonymous referee for the valuable suggestions and comments which led to the improvement of the manuscript.

References

[1] Abebe A., Singh P. (2008). An exact test for *U*-shaped alternatives of location parameters: The Exponential distribution case. *Communications in Statistics – Theory and Methods*, 37, 37-45.

[2] Bhat S.V. (2009). Some *k*-sample rank tests for umbrella alternatives. *Research Journal of Mathematics and Statistics*, 1(1), 27-29.

[3] Bhat S.V., Patil A.B. (2008). A class of *k*-sample distribution-free tests for the umbrella alternatives. *International Journal of Agricultural and Statistical Sciences*, 4(1), 53-61.

[4] Chen Y.I. (1993). Nonparametric comparisons of umbrella pattern treatment effects with a control in a one-way layout. *Communications in Statistics – Simulation and Computation*, 22(3), 749-764.

[5] Chen Y.I., Wolfe D.A. (1990). A study of distribution-free tests for umbrella alternatives. *Biometrical Journal*, 32, 47-57.

[6] Chen Y.I., Wolfe D.A. (1993). Nonparametric procedures for comparing umbrella pattern treatment effects with a control in a one-way layout. *Biometrics*, 49, 455-465.

[7] Deshpande J.V., Kusum K. (1984). A test for the non-parametric two sample scale problem. *Australian Journal of Statistics*, 26(1), 16-24.

[8] Gaur A., Mahajan, K.K., Arora S. (2012). New nonparametric tests for testing homogeneity of scale parameters against umbrella alternative. *Statistics and Probability Letters*, 82, 1681-1689.

[9] Gill A.N., Dhawan A.K. (1999). A one-sided test for testing homogeneity of scale parameters against ordered alternatives. *Communications in Statistics – Theory and Methods*, 28(10), 2417-2439.

[10] Kossler W. (2006). Some *c*-sample rank tests of homogeneity against umbrella alternatives with unknown peak. *Journal of Statistical Computation and Simulation*, 76(1), 57-74.

[11] Kusum K., Bagai I. (1988). A new class of distribution free procedures for testing homogeneity of scale parameters against ordered alternatives. *Communications in Statistics - Theory and Methods*, 17(5), 1365-1376.

[12] Lehmann E.L. (1963). Robust estimation in analysis of variance. *Annals of Mathematical Statistics*, 34, 957-966.

[13] Mack G.A., Wolfe D.A. (1981). *k*-sample rank tests for umbrella alternatives. *Journal of American Statistical Association*, 76, 175-181.

[14] Puri M.L. (1965). Some distribution-free *k*-sample rank order test of homogeneity against ordered alternatives. *Communication Pure Applied Mathematics*, 18, 51-63.

[15] Rao C.R. (1973). Linear Statistical Inference and its applications. *Wiley Eastern Ltd*, New York.

[16] Rao K.S.M. (1982). Non-parametric tests for homogeneity of scale against ordered alternatives. *Annals of Institute of statistical and Mathematics*, A(34), 327-334.

[17] Serfling R.J. (1980). Approximation theorems of mathematical statistics. *John Wiley*, New York.

[18] Shanubhogue A. (1988). Distribution-free test for homogeneity against stochastic ordering. *Metrika*, 35, 109-119.

[19] Shetty I.D., Bhat S.V. (1997). A distribution-free test for umbrella alternatives based on subsample medians. *Journal of Karnataka University Science*, 86-94.

[20] Simpson D.G., Margolin B.H. (1986). Recursive nonparametric testing for dose-response relationships subject to down-turns at high dose. *Biometrika*, 73, 589-596.

[21] Singh P., Gill A.N. (2004). A one-sided test based on sample quasi ranges. *Communications in Statistics – Theory and Methods*, 33(4), 835-849.

[22] Singh P., Liu W. (2006). A test against an umbrella ordered alternative. *Computational Statistics and Data Analysis*, 51, 1957-1964.