Statistical Methodology 29 (2016) 10–17

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

An adaptive test for the two-sample scale problem where the common quantile may be different from the median

W. Kössler [a,*], Narinder Kumar [b]

[a] Institute of Computer Science, Humboldt-Universität zu Berlin, Germany

[b] Department of Statistics, Panjab University, Chandigarh, India

ARTICLE INFO

Article history:
Received 12 October 2012
Received in revised form
18 August 2015
Accepted 20 August 2015
Available online 25 September 2015

Keywords:
Adaptive test
Asymptotic efficacy
Scale alternative
U-statistics
Tailweight

ABSTRACT

In the usual two-sample scale problem it is assumed that the two populations have a common median. We consider the case where the common quantile may be other than a half. We investigate a quite general class, all members are based on *U*-statistics where the minima and maxima of subsamples of various sizes are used. The asymptotic efficacies are investigated in detail. We construct an adaptive test where all statistics involved are suitably chosen. It is shown that the proposed adaptive test has good asymptotic and finite power properties.

1. Introduction

Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent random samples from a population with absolutely continuous distribution functions $F(x)$ and $F(x/e^{\vartheta})$. We wish to test

$$H_0 : \quad \vartheta = 0$$

against the alternative

$$H_1 : \quad \vartheta \neq 0.$$

* Correspondence to: Rudower Chaussee 25, D-12489 Berlin, Germany. Tel.: +49 3020933087.
E-mail address: koessler@informatik.hu-berlin.de (W. Kössler).

http://dx.doi.org/10.1016/j.stamet.2015.08.001
1572-3127/

The general two-sample scale problem was considered by Kochar and Gupta [6], Kössler [7,8], Hall et al. [4], Ramsey and Ramsey [18] and more recently, by Kössler and Kumar [10] and Marozzi [14,15].

Here, we consider the scale case when the populations have a common known quantile that may be of an order other than half. Such problems have many practical applications. As an example, cf. Deshpande and Kusum [1], let us have two filling machines that shall fill half of kg cans of dried milk. According to a given laid down criterion not more than five percent of the cans be under filled. Therefore both the machines are adjusted in such a way that five percent of the cans contain less than half kg and 95 percents can contain more than half kg of dried milk. In this case we can say that the distribution of the amount filled by the machines have a common quantile of order 0.05 and the more efficient machine is the one with smaller dispersion around this quantile.

Another field where such problem may occur is the pharmaceutical industry. Liu et al. [13] report on a drug–drug interaction study on a rheumatoid arthritis test drug where the test compound is taken by patients who are already under a particular medication (MTX, in this case). However the test compounds may result in a delayed elimination of MTX which is toxic if it remains too long in the body. Therefore the 24 h MTX plasma level should not exceed a certain threshold value t, say 5 picograms, with given probability q, e.g. $q = 0.9$. Assume now we have two test drugs under study, both satisfy the quantile condition $F(t) = q$ for the MTX level. Then the question is whether the MTX levels of the two test drugs differ in scale.

The ranked-set setting of such a question was considered in Öztürk and Deshpande [17] and in Gaur et al. [2].

Assume without restriction to the generality the common quantile is zero, i.e. $F(0) = \alpha$. Mehra and Rao [16] suggested the following kernel

$$\Phi(x_1,\ldots,x_k,y_1,\ldots,y_k) = \begin{cases} 1 & \text{if } \ 0 \le x_{(k)k} < y_{(k)k} \text{ and } 0 \le x_{(1)k}, y_{(1)k} \\ 1 & \text{if } \ y_{(1)k} < x_{(1)k} < 0 \text{ and } x_{(k)k}, y_{(k)k} < 0 \\ -1 & \text{if } \ 0 \le y_{(k)k} < x_{(k)k} \text{ and } 0 \le x_{(1)k}, y_{(1)k} \\ -1 & \text{if } \ x_{(1)k} < y_{(1)k} < 0 \text{ and } x_{(k)k}, y_{(k)k} < 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $x_{(i)k}$ is the ith order statistic in a subsample of size k from the X-sample (and likewise for y's). The suggested test statistic is

$$U_k = \frac{1}{\binom{n_1}{k}\cdot\binom{n_2}{k}} \sum \Phi(X_{r_1},\ldots,X_{r_k},Y_{s_1},\ldots,Y_{s_k}),$$

where the summation extends over all possible combinations $(r_1,\ldots,r_k)$ of k integers from $\{1,\ldots,n_1\}$ and all possible combinations $(s_1,\ldots,s_k)$ of k integers from $\{1,\ldots,n_2\}$. Obviously, large values of U_k speak for more scaled Y-sample.

Note that the kernel presented is a natural extension of Deshpande and Kusum [1] and Kusum [12].

The article is organized as follows. In Section 2 we give a rank representation of the test statistic that simplifies their computation considerably. In Section 3 we compute their asymptotic efficacies and intend to determine a suitable choice of subsample size k w.r.t. tail behavior of the underlying distribution. The results are used to define an adaptive test in Section 4. A simulation study is performed in Section 5. Section 6 gives a short example of the application of our test. Conclusions are drawn in Section 7.

2. Rank representation of U_k

Let us consider the positive and negative observations separately and let $n_j^<$ and $n_j^\ge, j = 1, 2$ the numbers of the negative and nonnegative observations, respectively. Let $R_{(s)}$ be the rank of the $Y_{(s)}$ observation in the pooled sample and $Q_{(s)} = R_{(s)} - n_1^< - n_2^<$. Assume first that we have no tied observations.

Then we may write U_k as

$$U_k = \frac{1}{\binom{n_1}{k}\cdot\binom{n_2}{k}} \Big((T_{1k} + T_{2k}) - (T_{3k} + T_{4k}) \Big),$$