Statistics and Probability Letters 119 (2016) 155–162

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Characterization based symmetry tests and their asymptotic efficiencies

B. Milošević, M. Obradović *

Faculty of Mathematics, University of Belgrade, Studenski trg 16, Belgrade, Serbia

A R T I C L E I N F O

Article history:
Received 22 December 2015
Received in revised form 6 July 2016
Accepted 7 July 2016
Available online 6 August 2016

MSC:
60F10
62F05
62G30

Keywords:
Symmetry tests
Order statistics
Bahadur efficiency
U-statistics

A B S T R A C T

A new characterization of symmetric distributions is presented. Two classes of distribution-free symmetry tests based on it are proposed. Their Bahadur efficiency is calculated and used for comparison with similar tests. Some classes of most favorable alternatives are also determined.

1. Introduction

The assumption of symmetry of distribution is important in many statistical procedures. Hence testing for symmetry has been a prominent topic in the statistics literature. Starting from the classical sign and Wilcoxon tests, a number of symmetry tests have been developed. Some of these can be found in e.g. Burgio and Nikitin (2001, 2007).

In recent times, introducing tests based on characterizations became a popular direction in goodness-of-fit testing. These kinds of tests, for different classes of distributions, can be found in, e.g. Henze and Meintanis (2002), Jovanović et al. (2015) and Volkova and Nikitin (2014). Symmetric distributions, as a special class of distributions, are no exception. Some symmetry tests based on characterizations have been studied in Baringhaus and Henze (1992), Nikitin (1996) and Litvinova (2001).

Ahsanullah (1992) proposed a characterization of symmetric distributions with respect to zero using the equidistribution of the absolute values of minimal and maximal order statistics from arbitrary sample size. Based on this characterization, Nikitin and Ahsanullah (2015) proposed two classes of symmetry tests. Here we generalize Ahsanullah's characterization for arbitrary order statistics and create new symmetry tests based on central order statistics.

2. Characterization and test statistics

Let $X_{(k;m)}$ be the kth order statistic from a sample of the size m. We present the following characterization of symmetry.

* Corresponding author.
E-mail address: marcone@matf.bg.ac.rs (M. Obradović).

http://dx.doi.org/10.1016/j.spl.2016.07.007
0167-7152/

Theorem 1. *Let $X_1, \dots, X_m$ be i.i.d. continuous random variables with distribution function $F(x)$ and let $k \le \frac{m}{2}$. Then the random variables $|X_{(k;m)}|$ and $|X_{(m-k+1;m)}|$ are equally distributed if and only if X_1 is symmetric with respect to zero.*

Proof. If X_1 is symmetric with respect to zero, then $|X_{(k;m)}|$ and $|X_{(m-k+1;m)}|$ are obviously equidistributed.

We now consider the "only if" part. Since with probability one $X_{(k;m)} < X_{(m-k+1;m)}$, from the equidistribution of $|X_{(k;m)}|$ and $|X_{(m-k+1;m)}|$ we have that $X_{(k;m)}$ and $-X_{(m-k+1;m)}$ have the same distribution, and the following identity holds

$$\sum_{j=k}^{m}\binom{m}{j}F(x)^j(1-F(x))^{m-j} = \sum_{j=k}^{m}\binom{m}{j}F(-x)^{m-j}(1-F(-x))^j. \tag{1}$$

Put

$$r(a) = \sum_{j=k}^{m}\binom{m}{j}a^j(1-a)^{m-j}.$$

Expression (1) then becomes $r(F(x)) = r(1-F(-x))$. Since $r(a)$ is the distribution function of the kth order statistic from the uniform distribution, it is monotonously increasing for $a \in (0,1)$ and we conclude that $F(x) = 1 - F(-x)$ for all $x \in R$, i.e. X_1 is symmetric. $\square$

Let $\mathcal{G}_0$ be the class of all continuous distributions that are symmetric about zero. Based on a sample of size n, $X_1, \dots, X_n$, from an unknown distribution F, we want to test the null hypothesis $H_0 : F \in \mathcal{G}_0$ against the alternative $H_1 : F \notin \mathcal{G}_0$. In view of our characterization, numerous possible tests can be constructed depending on the choice of k and m in Theorem 1. The tests based on extremal order statistics were considered in Nikitin and Ahsanullah (2015). Here we propose two classes of tests, of integral and Kolmogorov type, based on the equidistribution of two central order statistics from the subsamples of even size $m = 2k$.

These tests have the following test statistics

$$J_n^k = \int_0^\infty (H_n^{(k)}(t) - G_n^{(k)}(t))dQ_n(t), \tag{2}$$

$$K_n^k = \sup_{t>0}\left|H_n^{(k)}(t) - G_n^{(k)}(t)\right|, \tag{3}$$

where Q_n is empirical distribution function of the sample $|X_1|, \dots, |X_n|$ and

$$H_n^{(k)}(t) = \frac{1}{\binom{n}{2k}}\sum_{\mathcal{I}_{2k}} I\{|X_{(k),X_{i_1},\dots,X_{i_{2k}}}| < t\}, \tag{4}$$

$$G_n^{(k)}(t) = \frac{1}{\binom{n}{2k}}\sum_{\mathcal{I}_{2k}} I\{|X_{(k+1),X_{i_1},\dots,X_{i_{2k}}}| < t\}, \tag{5}$$

are U-empirical distribution functions related to the characterization. Here $X_{(k),X_1,\dots,X_m}$ denotes the kth order statistic from the sample $X_1, \dots, X_m$ and $\mathcal{I}_m = \{(i_1, \dots, i_m) : 1 \le i_1 < \cdots < i_m \le n\}$. We consider large values of both statistics to be significant. The tests with test statistics J_n^k are not consistent against all alternatives, however, consistency takes place for many common alternatives.

Next we show that our test statistics are distribution-free under the null hypothesis of symmetry. Statistic J_n^k can be written as

$$J_n^k = \int_{\frac{1}{2}}^1 (H_n^{(k)}(F^{-1}(y)) - G_n^{(k)}(F^{-1}(y)))dQ_n(F^{-1}(y)), \tag{6}$$

where $F^{-1}(y)$ is the inverse of the distribution function, assuming, for simplicity, that it is strictly monotone. For $y > \frac{1}{2}$ we have

$$H_n^{(k)}(F^{-1}(y)) = \frac{1}{\binom{n}{2k}}\sum_{\mathcal{I}_{2k}} I\{|X_{(k),X_{i_1},\dots,X_{i_{2k}}}| < F^{-1}(y)\}.$$

Using the symmetry of the null distribution and the probability transform, we get

$$H_n^{(k)}(F^{-1}(y)) = \frac{1}{\binom{n}{2k}}\sum_{\mathcal{I}_{2k}} I\{1-y < U_{(k),U_{i_1},\dots,U_{i_{2k}}} < y\},$$

where $U_1, \dots, U_n$ are independent random variables with uniform $\mathcal{U}[0,1]$ distribution.

In a similar manner, both $G_n^{(k)}(F^{-1}(y))$ and $Q_n^{(k)}(F^{-1}(y))$ can also be expressed as functions of $U_1, \dots, U_n$. Thus, the statistics J_n^k are distribution free. Analogously, the same holds for the statistics K_n^k.

Therefore without loss of generality we may suppose that, under the null hypothesis, the random variables $X_1, \dots, X_n$ are from the uniform $\mathcal{U}[-1,1]$ distribution.