COMSTA: 6610 Model 3G pp. 1–17 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Testing for central symmetry and inference of the unknown center

Xinjie Dai [a], Cuizhen Niu [b], Xu Guo [b,*]

[a] Department of Astronomy, Beijing Normal University, Beijing, PR China
[b] School of Statistics, Beijing Normal University, Beijing, PR China

A R T I C L E I N F O

Article history:
Received 20 August 2017
Received in revised form 26 April 2018
Accepted 9 May 2018
Available online xxxx

Keywords:
Characteristic function
Hodges–Lehmann estimator
Symmetry
Testing

A B S T R A C T

In this paper, we consider testing for central symmetry and inference of the unknown center with multivariate data. Our proposed test statistics are based on weighted integrals of empirical characteristic functions. With two special weight functions, we obtain test statistics with simple and closed forms. The test statistics are easy to implement. In fact, they are based merely on pairwise distances between points in the sample. The asymptotic results are developed. It is proven that our proposed tests can converge to finite limit at the rate of n^{-1} under the null hypothesis and can detect any fixed alternatives. For the unknown center, we also propose two classes of minimum distance estimators based on the previously introduced test statistics. The asymptotic normalities are derived. Efficient algorithms are also developed to compute the estimators in practice. We further consider checking whether the unknown center is equal to a specified value μ_0. Extensive simulation studies and one medical data analysis are conducted to illustrate the merits of the proposed methods.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

In this paper, we consider testing for central symmetry and inference of the unknown center with multivariate data. These problems are very important and classical in statistics, with applications ranging from ecology, environment to economics and finance (see Laska et al., 2014; Lyubchich et al., 2016; Henderson and Parmeter, 2015, and references therein). Suppose $X \in \mathbb{R}^p$ be a random vector. The hypothesis of interest is as follows:

$$H_0 : X - \mu \overset{d}{=} \mu - X,$$

where μ is some unknown center and $\overset{d}{=}$ stands for equality in law. The above introduced symmetry is interchangeably called central symmetry (Einmahl and Gan, 2016), reflected symmetry (Henze et al., 2003) or diagonal symmetry (Székely and Sen, 2002), respectively. We also consider estimating the unknown center of a multivariate symmetric distribution. Further, we investigate how to check whether the unknown center μ is a specified value μ_0.

Many authors have studied central symmetry testing problem. For univariate data, see for instance, Fang et al. (2015), Henderson and Parmeter (2015), Laska et al. (2014), and Lyubchich et al. (2016). For multivariate data, among others, Székely and Móri (2001) found that central symmetry around zero holds if and only if $E(\|X+X'\|) = E(\|X-X'\|)$. Here $\|\cdot\|$ is the Euclidean distance and X, X' follow independent and identical distribution (i.i.d.). Based on empirical characteristic function, Henze et al. (2003) developed a flexible class of omnibus affine invariant tests for central symmetry. A similar approach was

* Corresponding author.
E-mail address: xguo12@bnu.edu.cn (X. Guo).

https://doi.org/10.1016/j.csda.2018.05.007
0167-9473/© 2018 Elsevier B.V. All rights reserved.

Please cite this article in press as: Dai X., et al., Testing for central symmetry and inference of the unknown center. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.05.007.

UNCORRECTED PROOF

also adopted by Neuhaus and Zhu (1998). Ngatchou-Wandji (2009) and Quessy (2016) made some further extensions of Henze et al. (2003)'s method.

Another important problem is to estimate the unknown center of a symmetric distribution. A simple choice is the sample average. However, the sample average can perform poorly when there are outliers or the data exhibits heavy tails. On the other hand, although sample median can be robust, it is not very efficient. Then it is important and interesting to obtain robust and efficient estimators of the unknown center. For $p = 1$, the Hodges–Lehmann estimator (Hodges and Lehmann, 1963) and the location M-estimator of Huber (1964) are two popular estimators. Recently, based on the characteristic function, Bondell (2008) proposed a class of alternative robust and efficient estimators of the unknown center. For $p \geq 1$, Chaudhuri (1992) proposed multivariate extension of Hodges–Lehmann estimator.

In this paper, we generalize some existing methods about testing for central symmetry. To be precise, the test statistics introduced by Székely and Móri (2001) and Henze et al. (2003) are revisited and generalized. The asymptotic properties are also developed by using the theory of V-statistics, which makes the proofs easier to access. It is proven that our proposed test statistics can converge to finite limit at the rate of n^{-1} under the null hypothesis and can detect any fixed alternatives. Afterwards, we propose two classes of minimum distance estimators based on the previously introduced test statistics for the unknown center. They can be viewed as generalizations of the methods in Bondell (2008) from univariate data to multivariate data. The asymptotic normalities are derived. Efficient algorithms are also developed to compute the estimators in practice. We further consider checking whether the unknown center is equal to a specified value μ_0.

The proposed test statistics are easy to implement. In fact, they are based on interpoint distances. Recently, interpoint distances based tests are found very useful for the two-sample comparison problem, especially in high dimensional situation. Jureckova and Kalina (2012) introduced some rank tests based on interpoint distances. Marozzi (2016) modified Jureckova and Kalina's tests by a more efficient use of the data. Noting that different tests should have different power performances for the same distribution function, Marozzi (2015) further developed combined multi-distance tests. For other recent developments, see also Baringhaus and Franz (2010), Székely and Rizzo (2013), and Biswas and Ghosh (2014).

The rest of this paper is organized as follows. In Section 2, we construct the test statistics for central symmetry. We derive their asymptotic properties under null hypothesis and fixed alternative hypotheses in this Section. Inference procedures for the unknown center are developed also in Section 2. In Section 3, some extensive simulation analyses are carried out to illustrate the merits of the proposed procedures. One medical data analysis is given in Section 4. Some conclusions and discussions are put in Section 5. All proofs of the theoretical results are postponed to the Appendix.

2. Testing and estimation

In this section, we propose methods to test for central symmetry and also develop procedures to infer about the unknown center. First, we consider the central symmetry checking problem.

2.1. Testing for central symmetry

Denote $Y = X - \mu$. The null hypothesis then can be rewritten as $H_0: \; Y \overset{d}{=} -Y$. This is further equivalent to

$$H_0: \; \phi_Y(t) = E(e^{it^\tau Y}) = E(e^{-it^\tau Y}) = \phi_{-Y}(t), \;\; \forall t \in \mathbb{R}^p.$$

Here we use $\phi_Y(t)$ and $\phi_{-Y}(t)$ to denote the characteristic functions of Y and $-Y$, respectively. This motivates us to consider the following quantity:

$$D_\omega = \int_{\mathbb{R}^p} |\phi_Y(t) - \phi_{-Y}(t)|^2 \omega(t) dt. \qquad (1)$$

Here $|c_1 + ic_2| = \sqrt{c_1^2 + c_2^2}$ is the modulus of the complex number $c_1 + ic_2$.

Note that:

$$|\phi_Y(t) - \phi_{-Y}(t)|^2 = 2E\left[\cos(t^\tau(Y - Y')) - \cos(t^\tau(Y + Y'))\right],$$

with Y' being independent copy of Y. We first make use of the fact that the characteristic function of a spherical stable law is given by

$$\phi_Z(t) = \int_{\mathbb{R}^p} \cos(t^\tau z) f_{a,p}(z) dz = e^{-\|t\|^a}.$$

Here $f_{a,p}(\cdot)$ stands for the density of a spherical stable law in $\mathbb{R}^p$ with characteristic exponent $a \in (0, 2]$.

From Samorodnitsky and Taqqu (1994) and also Nolan (2013), we know that if Z is a-stable and elliptically contoured, it has characteristic function:

$$E \exp(it^\tau Z) = \exp(-(t^\tau Q t)^{a/2} + it^\tau \delta),$$

for some $p \times p$ positive definite matrix Q and shift vector $\delta \in \mathbb{R}^p$. The spherical symmetric cases arise when Q is a multiple of the identity matrix and $\delta = 0$, in which case the characteristic function simplifies to

$$E \exp(it^\tau Z) = \exp(-\gamma^a \|t\|^a).$$

If we further set $\gamma = 1$, standard spherical stable law arises.

UNCORRECTED PROOF

Please cite this article in press as: Dai X., et al., Testing for central symmetry and inference of the unknown center. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.05.007.

Download English Version:

https://daneshyari.com/en/article/6868622

Download Persian Version:

https://daneshyari.com/article/6868622

Daneshyari.com