Journal of the Korean Statistical Society 44 (2015) 577–591

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Nonparametric homogeneity test based on ridit reliability functional

Uttam Bandyopadhyay *, Debajit Chatterjee

Department of Statistics, University of Calcutta, 35 Ballygunge Circular Road, Kolkata – 700019, India

ARTICLE INFO

Article history:
Received 19 May 2014
Accepted 7 March 2015
Available online 30 March 2015

AMS 2000 subject classifications:
primary 62G10
secondary 62N05

Keywords:
Asymptotic distribution
Nonparametric homogeneity problem
Ridit
U-statistic

ABSTRACT

The paper provides nonparametric test procedures for comparing K (>2) unknown univariate populations, in which the tests are formulated by using consistent estimators of the ridit reliability functionals (see, for example, Bandyopadhyay and De, 2011) for comparing more than two populations. The tests are asymptotically distribution free and can be used for data on both discrete and continuous random variables. An extensive numerical study is performed to compare the proposed test with the nonparametric tests, obtained from Konietschke et al. (2012), in terms of type I error rate and power. A data study illustrates the use of such tests.

1. Introduction

There is an extensive literature on the analysis of data corresponding to two or more independent samples (see, for example, Camblor & Alvarev, 2009; Nishiyama, 2011; Zhang, 2006; Zhang & Wu, 2007). The present work is related to nonparametric approach without assuming that the underlying distribution functions are continuous. Under nonparametric approach, the test by Brunner and Munzel (2000) is framed through the ridit reliability functional (see, for example, Beder & Heim, 1990)

$$\tau = P(X_1 < X_2) + \frac{1}{2}P(X_1 = X_2),$$

where the random variables X_1 and X_2 have the common support on the real line and are distributed independently according to the distribution functions F_1 and F_2, respectively. The interpretation of τ as a ridit reliability functional is as follows: The subjects under F_2 will be called more (or less) reliable than the subjects under F_1 if $P(X_1 < X_2) >$ (or $<$)$P(X_1 > X_2)$, which, after simplification, is equivalent to $\tau >$ (or $<$)$\frac{1}{2}$. Furthermore, subjects under F_2 will be considered as reliable as that under F_1 if $\tau = \frac{1}{2}$. Now the quantity τ can be expressed as

$$\tau = \int \left[P(X_1 < x) + \frac{1}{2}P(X_1 = x) \right] dF_2(x),$$

in which the quantity

$$P(X_1 < x) + \frac{1}{2}P(X_1 = x)$$

* Corresponding author.
E-mail addresses: ubstat@caluniv.ac.in (U. Bandyopadhyay), debajit1.chatterjee@gmail.com (D. Chatterjee).

http://dx.doi.org/10.1016/j.jkss.2015.03.004
1226-3192/

can be interpreted as the ridit score for F_1 at the realized value x in the sample space. Thus τ represents the average ridit of F_1 relative to F_2, and hence defined as the ridit reliability functional. The use of such functional can be found in the following classical testing problems:

(i) Let X_k have normal distribution with unknown mean (μ_k) and unknown variance (σ_k^2), $k = 1, 2$. Then, writing $\Phi(.)$ as the distribution function of the standard normal variable, it is not difficult to get

$$\tau = P(X_1 < X_2) = \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right).$$

(ii) Let X_k follow Bernoulli distribution with unknown mean (μ_k), $0 < \mu_k < 1$, $k = 1, 2$. Then

$$\tau = \frac{1}{2} + \frac{1}{2}(\mu_2 - \mu_1).$$

(iii) Let X_k follow geometric distribution with unknown mean (μ_k^{-1}), $0 < \mu_k < \infty$, $k = 1, 2$. Then, a simple algebraic manipulation yields

$$\tau = \frac{1}{2} + \frac{1}{2}\frac{(\mu_2 - \mu_1)}{(\mu_1 + \mu_2 + \mu_1\mu_2)}.$$

In all the above situations, it can be seen that

$$\tau \gtreqless \frac{1}{2} \text{ as } \mu_2 \gtreqless \mu_1.$$

Furthermore, the works by Bandyopadhyay and Biswas (2001) and Bandyopadhyay and De (2007) provide applications on τ in sequential design. In addition to the parametric models (i)–(iii), it is possible to get $\tau = \frac{1}{2}$ when F_1 and F_2 are both continuous with $F_1 = F_2$ or with $F_2(x) = F_1(\frac{x}{\sigma})$, $\sigma > 0$ and $F_1(x) = F_1(-x)$ for all x. Specifically, when F_1 and F_2 are both continuous, τ becomes Mann–Whitney functional. With this background, the K-sample homogeneity problem is revisited and some nonparametric tests are formulated using a set of K functionals that generalizes τ.

A kind of generalization of τ corresponding to K-sample Behrens–Fisher problem is given by the set of functionals (see Akritas, Arnold, & Brunner, 1997; Brunner & Puri, 2001; Konietschke, Hothorn, & Brunner, 2012)

$$\tau_k = P(X_G < X_k) + \frac{1}{2}P(X_G = X_k), \quad k = 1, 2, \ldots, K, \tag{1.1}$$

where $X_k \sim F_k$, $k = 1, 2, \ldots, K$ independently and $X_G \sim G = \sum_{k=1}^{K} w_k F_k$, $0 \le w_k \le 1$, $k = 1, 2, \ldots, K$ with $\sum_{k=1}^{K} w_k = 1$. For K-sample location problem, the use of alternative hypothesis dependent single ridit reliability functional can be found in the work of Terpstra and Magel (2003). For example, the functional with K variables $X_1, X_2, \ldots, X_K$

$$P(X_1 \le X_2 \le \cdots \le X_K \text{ with strict inequality for at least one pair})$$

is used for ordered relation of K locations. In the present work, τ is generalized in the line of Bandyopadhyay and De (2011), where a set of ridit reliability functionals is used to compare Bernoulli success probabilities. Such functionals are considered to construct some tests for general set up.

The content of the paper is as follows. The set up and the homogeneity hypothesis are described in Section 2. Section 3 contains the proposed asymptotically distribution free (ADF) tests. Section 4 provides simulated values of type I error rate and power for various test procedures, proposed and competitors. A data study to illustrate the use of the different test procedures is given in Section 5. The paper concludes in Section 6, followed by some technical details in the Appendices.

2. Setup and hypothesis

Consider a set of K independent samples, in which the kth sample corresponds to the random variable X_k having the distribution function F_k, $k = 1, 2, \ldots, K$. The samples are denoted by

$$X_k : X_{k1}, X_{k2}, \ldots, X_{kn_k}, \quad k = 1, 2, \ldots, K.$$

The object of the present work is to formulate nonparametric procedures for testing the null hypothesis H_0: $\left[F_1 = F_2 = \cdots = F_K\right]$ against an alternative H_a without assuming that F_k is continuous, $k = 1, 2, \ldots, K$. The problem is known as nonparametric K-sample homogeneity problem. As mentioned earlier, the reliability functionals (see Bandyopadhyay & De, 2011), given by

$$\begin{aligned} R_k &= P[X_k > \max(X_m, m = 1, 2, \ldots, K; m \ne k)] \\ &\quad + \frac{1}{2}\sum_{1 \le k_1 \le K, k_1 \ne k} P[X_k = X_{k_1} > \max(X_m, m = 1, 2, \ldots, K; m \ne k, k_1)] \end{aligned}$$