Journal of Multivariate Analysis 102 (2011) 629–640

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Restricted one way analysis of variance using the empirical likelihood ratio test

Wen Yu [a], Hammou El Barmi [b,*], Zhiliang Ying [c]

[a] Department of Statistics, School of Management, Fudan University, Shanghai 200433, PR China
[b] Department of Statistics and Computer Information Systems, Baruch College, The City University of New York, New York, NY 10010, United States
[c] Department of Statistics, Columbia University, New York, NY 10027, United States

ARTICLE INFO

Article history:
Received 26 December 2008
Available online 25 November 2010

AMS subject classifications:
62E20

Keywords:
One way analysis of variance
Isotonic regression
Empirical likelihood ratio test
Chi-bar squared distribution

ABSTRACT

Empirical likelihood (EL) ratio tests are developed for testing for or against the hypothesis that k-population means $\mu_1, \mu_2, \ldots, \mu_k$ are isotonic with respect to some quasi-order $\preceq$ on $\{1, 2, \ldots, k\}$. The null asymptotic distributions are derived and are shown to be of chi-bar squared type. The asymptotic power of the proposed test for testing for equality of these means against the order restriction is derived under contiguous alternatives and a simulation study is carried out to investigate the finite sample behaviors of this test. In addition, an adjusted EL test is used to improve the small size performance of our test and an example is also discussed to illustrate the theoretical results.
© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Many problems in statistics are concerned with comparing the means of several populations based on independent random samples. For example, in agricultural experiments, it is often of interest to compare the average crop yields under different conditions and in reliability studies, investigators may be interested in comparing the mean lifetimes under different stress conditions. In general, one assumes that $\mu_1, \mu_2, \ldots, \mu_k$ are means corresponding to k-populations and considers testing $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ where

$$\mathcal{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad \text{and} \quad \mathcal{H}_2 : \text{No constraints on } (\mu_1, \mu_2, \ldots, \mu_k)^T. \tag{1}$$

Many tests have been developed for this situation and they include one way analysis of variance F-test and nonparametric tests based on ranks. These tests are not designed to detect any particular departure from $\mathcal{H}_0$.

In some instances, the population means are believed to be isotonic with respect to some quasi-order $\preceq$ on $S = \{1, 2, \ldots, k\}$. For example, it is reasonable to assume that the averages of the times required for successful completion of physical therapy for individuals who undergo a knee surgery are ordered according to prior physical fitness status which can be categorized into below average, average and above average. When such kind of prior knowledge is available, it could be utilized by restricting the parameter space in the alternative hypothesis to $\mathcal{H}_1 - \mathcal{H}_0$ where

$$\mathcal{H}_1 : (\mu_1, \mu_2, \ldots, \mu_k)^T \text{ is isotonic with respect to } \preceq. \tag{2}$$

When $\preceq$ corresponds to the simple order, the likelihood ratio test for this situation in the normal case when the variance are known or are unknown but equal was studied in [1]. More general situations that include the case of unknown and unequal variances were considered in [11]. We note that some tests designed for testing $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ have been extended

* Corresponding author.
E-mail address: hammou_elbarmi@baruch.cuny.edu (H. El Barmi).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.11.006

to this situation and incorporating such information has been shown to improve statistical inferences. For more on this, see [15,18].

Owen [7,8] introduced the empirical likelihood (EL) approach to statistical inference. He used it to construct generalized likelihood ratio test statistics and corresponding confidence regions. He also showed that a nonparametric version of Wilks' theorem [19] holds when the EL is used. The advantage of this nonparametric approach is that it allows one to consider a very large class of distributions without having to specify a particular parametric model. This procedure has sampling properties similar to the bootstrap. However, where the bootstrap uses resampling, the EL profiles a multinomial likelihood supported on the sample. It is also very similar to that used in parametric models but is more complex in terms of computations. The approach has since been extended to many areas in statistics and many references can be found in [10].

It is well known that the likelihood function provides an efficient way to extract information from data through statistical models [12]. It is also known that the EL based tests are parameter transformation invariant and Bartlett correctable [10,4]. As a byproduct of the nonparametric estimation of the underlying probability distribution, the EL provides estimation of the underlying cumulative distributions under various hypotheses (constraints). Such nonparametric estimates can be made use of, in some situations, including the case-control logistic regression, to test goodness of fit [14].

Owen [9] developed an EL test to test $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ and showed that the limiting distribution of this test is χ^2_{k-1}. El Barmi [5] extended the EL approach to situations where order constraints are present. Specifically, he assumed that $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{ki})^T,\ i = 1, 2, \ldots, n$, is a random sample from a population whose distribution function depends on a $k \times 1$ parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$ through $E_F \psi(\mathbf{X}, \boldsymbol{\theta}) = 0$ where ψ is a known k-dimensional function. Based on this random sample, he tested $\mathcal{H}_0^* : g_i(\boldsymbol{\theta}) = 0,\ i = 1, 2, \ldots, s$, against $\mathcal{H}_1^* - \mathcal{H}_0^*$ where $\mathcal{H}_1^* : g_i(\boldsymbol{\theta}) \le 0,\ i = 1, 2, \ldots, s$, and g_i is a well behaved real valued function for all $i \le s$ and $s \le k$. His test statistic is given by

$$T^*_{01} = -2 \log \frac{\max\left\{\prod_{i=1}^n p_i, \sum_{i=1}^n p_i \psi(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{0}, \sum_{i=1}^n p_i = 1, g_i(\boldsymbol{\theta}) = 0,\ i = 1, 2, \ldots, s\right\}}{\max\left\{\prod_{i=1}^n p_i, \sum_{i=1}^n p_i \psi(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{0}, \sum_{i=1}^n p_i = 1, g_i(\boldsymbol{\theta}) \le 0,\ i = 1, 2, \ldots, s\right\}}.$$

He showed that the null limiting distribution of T^*_{01} is of chi-bar squared type under some regularity conditions and gave the expression of the weighting values. As a consequence of his results and under this setting, if one takes $\psi(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{X} - \boldsymbol{\theta}$ so that $\boldsymbol{\theta} = \boldsymbol{\mu}$ and chooses g_is appropriately to correspond to $\mathcal{H}_1$, then the null limiting distribution of the EL ratio test statistic for testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ is of a chi-bar squared type.

A situation commonly encountered in practice is, given $\{X_{i1}, X_{i2}, \ldots, X_{in_i}\},\ i = 1, 2, \ldots, k$, independent random samples from the k-populations, test $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ where $\mathcal{H}_0$ and $\mathcal{H}_1$ are as defined in (1) and (2). Our goal here is to extend existing results to this situation and obtain the corresponding EL ratio test and its asymptotic distribution. We also test $\mathcal{H}_1$ against $\mathcal{H}_2 - \mathcal{H}_1$ and show that the results obtained under normal theory when the variances are known hold asymptotically here. In addition, we obtain the expression of the asymptotic power of our test for testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ under contiguous alternatives using LeCam's third lemma.

It is well known that the estimated error rates from the unconstrained EL ratio tests using a χ^2 threshold could deviate significantly from the corresponding nominal levels for small sizes. Various proposals to reduce such large deviations have been discussed in the literature and they include the adjusted EL introduced by Chen et al. [2]. This technique will also be adopted here in order to improve the estimation of the error rates of our tests.

The rest of paper is organized as followed. In Section 2, we briefly outline Owen's EL based one way analysis of variance. In Section 3, the EL ratio tests for testing $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ and $\mathcal{H}_1$ against $\mathcal{H}_2 - \mathcal{H}_1$ are developed and their limiting distributions under $\mathcal{H}_0$ are obtained. We also explore the asymptotic power of the proposed test for $\mathcal{H}_0$ against $\mathcal{H}_1 - \mathcal{H}_0$ under contiguous alternatives. In Section 4, adjusted versions of the proposed EL ratio tests are introduced in order to improve the small sample behavior of our tests. In Section 5, some simulation studies are carried out to investigate the finite sample performance of the proposed tests and a real example is used to illustrate the theoretical results. Section 6 gives some concluding remarks. All the technical details are given in the Appendix. Throughout we use $\mathcal{H}_i$ to also denote the region in $\mathbb{R}^k$ defined by the hypothesis $\mathcal{H}_i,\ i = 0, 1, 2$ and by convention, $\sup_\emptyset = 0$ and $0/0 = 1$. Finally, we note that Davidov et al. [3] have developed an empirical likelihood based test for testing for or against likelihood ratio ordering among several populations whose ratios of probability densities are of a particular parametric form.

2. Empirical likelihood based one way analysis of variance

Let $\{X_{i1}, X_{i2}, \ldots, X_{in_i}\},\ i = 1, 2, \ldots, k$, be independent random samples from k-populations with means $\mu_1, \mu_2, \ldots, \mu_k$, respectively. We assume that their corresponding distributions, $F_1, F_2, \ldots, F_k$, are unknown and wish to test $\mathcal{H}_0$ against $\mathcal{H}_2 - \mathcal{H}_0$ where these hypotheses are as defined before. Since a parametric likelihood is not available, we use instead the nonparametric likelihood

$$\prod_{i=1}^k \prod_{j=1}^{n_i} p_{ij} \tag{3}$$

where $p_{ij} \ge 0$ for all (i, j) and $\sum_{j=1}^{n_i} p_{ij} = 1,\ i = 1, 2, \ldots, k$.

Download English Version:

https://daneshyari.com/en/article/1146054

Download Persian Version:

https://daneshyari.com/article/1146054

Daneshyari.com