Computational Statistics and Data Analysis 93 (2016) 285–293

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Empirical likelihood confidence regions for one- or two-samples with doubly censored data

Junshan Shen [a], Kam Chuen Yuen [b], Chunling Liu [c,*]

[a] School of Mathematical Sciences, Peking University, Yiheyuan Road, Beijing, PR China

[b] Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong

[c] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 30 April 2014
Received in revised form 18 January 2015
Accepted 18 January 2015
Available online 28 January 2015

Keywords:
Chi-square convergence
Confidence region
Doubly censored data
EM algorithm
Empirical likelihood ratio
Moment constraint

ABSTRACT

The purpose is to propose a new EM algorithm for doubly censored data subject to non-parametric moment constraints. Empirical likelihood confidence regions are constructed for one- or two- samples of doubly censored data. It is shown that the corresponding empirical likelihood ratio converges to a standard chi-square random variable. Simulations are carried out to assess the finite-sample performance of the proposed method. For illustration purpose, the proposed method is applied to a real data set with two samples.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The empirical likelihood (EL) method of Owen (1988) is a very useful tool for constructing confidence regions in non-parametric problems. Since it has many desirable statistical properties and some advantages over other competitors such as the normal approximation and the bootstrap, much research on the EL method has been carried out by many authors. Empirical likelihood and empirical entropy are two different generalizations of the likelihood method in the Cressie–Read family. For the latter, Lee and Oh (2015) derived an entropy based test for autoregressive conditional duration (ACD) models. For the former, DiCiccio et al. (1991) proved that the EL method is Bartlett-correctable in a very general setting; Chen (1993) investigated the coverage error of EL confidence regions for a linear regression model; and Chen and Hall (1993) studied coverage accuracy of smoothed EL confidence regions for quantiles; and Qin and Lawless (1994) developed the EL method to deal with the general estimating equations, among others. See Owen (2001) for a comprehensive review.

In recent years, the EL method has been employed to deal with incomplete data. One way to do it is to construct synthetic complete data from incomplete sample. Using this approach, Wang and Jing (2001) constructed the EL confidence intervals for linear functional of survival function for right censored data; and Wang and Rao (2002) used the EL method to make inference for the mean of response variable under kernel regression imputation for missing response data. Unfortunately, since the synthetic data are not independent and identically distributed, the corresponding EL ratio no longer converges to a standard chi-square random variable, and hence one needs to modify the EL ratio by means of an adjustment coefficient which is sometimes very difficult to estimate. Another approach to handling incomplete sample is to construct likelihood

* Corresponding author. Tel.: +852 2766 6931; fax: +852 2362 9045.
E-mail address: macliu@polyu.edu.hk (C. Liu).

http://dx.doi.org/10.1016/j.csda.2015.01.010
0167-9473/© 2015 Elsevier B.V. All rights reserved.

ratio directly from incomplete data. This method can be traced back to Thomas and Grunkemeier (1975) who used the nonparametric likelihood ratio method to construct confidence intervals for survival probabilities under random censorship (see also Li (1995)). For right-censored data, Li and Van Keilegom (2002) obtained confidence bands for conditional survival and quantile functions using the nonparametric likelihood ratio approach. Murphy and van der Vaart (1997) extended the EL method to semiparametric models and showed that the likelihood ratios for the models satisfy the Wilks theorem. As the likelihood ratios for semiparametric models are much more complex than the saturated nonparametric model, one needs to develop a computational algorithm to make the theoretical results practical. By introducing a hidden weight for each subject, Zhou (2005b) proposed a modified EM algorithm to calculate the EL ratio for censored or truncated data. Using this method, Zhou (2005a) performed EL analysis of the rank estimator for the censored accelerated failure time model.

The self-consistent algorithm was first used in Efron (1967) to calculate the Kaplan–Meier estimator. Turnbull (1974) and Mykland and Ren (1996) investigated self-consistent estimators for doubly censored data. In this paper, we propose a new self-consistent EM algorithm with constraint to deal with the likelihood ratio for doubly censored data. Compared to the method of Zhou (2005b), the proposed algorithm does not need to introduce hidden weights, and hence can be easily implemented in practice.

The rest of this paper is organized as follows. In Section 2, we propose a self-consistent algorithm to calculate the EL for one-sample problem with doubly censored data, and study some simple properties of the algorithm. In Section 3, we show how to apply the algorithm to two-sample problem with doubly censored data. In Section 4, we carry out simulation studies to assess the performance of the proposed method, and apply the algorithm to a real data set. Finally, a brief discussion about the proposed method is given in Section 5, and the proof of the main result is presented in the Appendix.

2. One-sample problem with doubly censored data

We first review the basic idea of empirical likelihood for completely observed data. Let X be a random variable with distribution F satisfying

$$E_F(\mathbf{m}(X,\boldsymbol{\theta})) = \mathbf{0}, \tag{1}$$

where $\boldsymbol{\theta}$ is a p-dimensional parameter vector of interest and $\mathbf{m}(x,\boldsymbol{\theta})$ is a r-dimensional ($r \geq p$) vector. The nonparametric likelihood based on the random sample $\{X_i, i = 1,\ldots,n\}$ can be written as

$$L(F) = \prod_{i=1}^{n} \Delta F(X_i),$$

where $\Delta F(x) = F(x) - F(x-)$. For a given $\boldsymbol{\theta}$, define the profile likelihood

$$\mathcal{L}(\boldsymbol{\theta}) = \sup_F \left\{ L(F) | E_F(\mathbf{m}(X,\boldsymbol{\theta})) = \mathbf{0} \right\}.$$

Let $p_i = P(X = X_i)$ be the probability weight for datum X_i ($i = 1,\ldots,n$) and define $p = (p_1,\ldots,p_n)^T$. Then, the likelihood $\mathcal{L}(\boldsymbol{\theta})$ becomes

$$\mathcal{L}(\boldsymbol{\theta}) = \sup_p \left\{ \prod_{i=1}^{n} p_i \middle| \sum_{i=1}^{n} p_i \mathbf{m}(X_i,\boldsymbol{\theta}) = \mathbf{0}, \sum_{i=1}^{n} p_i = 1, p_i \geq 0 \right\}.$$

Define the Lagrange function

$$l(\boldsymbol{\theta}, p, \boldsymbol{\lambda}, \gamma) = \sum_{i=1}^{n} \log p_i - n\boldsymbol{\lambda}^T \sum_{i=1}^{n} p_i \mathbf{m}(X_i,\boldsymbol{\theta}) + \gamma \left(\sum_{i=1}^{n} p_i - 1 \right),$$

where $\boldsymbol{\lambda}$ and γ are the Lagrange multipliers. By the Lagrange method, $L(F)$ attains its maximum value under the constraint $E_F(\mathbf{m}(X,\boldsymbol{\theta})) = \mathbf{0}$ at critical points of $l(\boldsymbol{\theta}, p, \boldsymbol{\lambda}, \gamma)$

$$p_i = \frac{1}{n} \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{m}(X_i,\boldsymbol{\theta})}, \quad i = 1,\ldots,n,$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta})$ satisfies

$$\sum_{i=1}^{n} \frac{\mathbf{m}(X_i,\boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^T \mathbf{m}(X_i,\boldsymbol{\theta})} = \mathbf{0}.$$

So, the logarithm of $\mathcal{L}(\boldsymbol{\theta})$ becomes

$$\log \mathcal{L}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \log\left(1 + \boldsymbol{\lambda}^T \mathbf{m}(X_i,\boldsymbol{\theta})\right) - n \log n.$$

Download English Version:

https://daneshyari.com/en/article/415329

Download Persian Version:

https://daneshyari.com/article/415329

Daneshyari.com