Journal of Statistical Planning and Inference 146 (2014) 95–101

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Smoothed empirical likelihood inference for the difference of two quantiles with right censoring

Hanfang Yang [a], Crystal Yau [b], Yichuan Zhao [c,*]

[a] School of Statistics, Renmin University of China, Beijing, China
[b] Transmission Financial and Accounting Controls, Georgia Power, Atlanta, GA 30308, United States
[c] Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, United States

ARTICLE INFO

Article history:
Received 9 March 2012
Received in revised form
26 June 2013
Accepted 20 September 2013
Available online 14 October 2013

Keywords:
Censored data
Difference of two quantiles
Smoothed empirical likelihood
Confidence interval

ABSTRACT

In this paper, using a smoothed empirical likelihood method, we investigate the difference of quantiles in the two independent samples and construct the confidence intervals. We prove that the limiting distribution of the empirical log-likelihood ratio is a chi-squared distribution like Shen and He (2007). In the simulation studies, in terms of coverage accuracy and average length of confidence intervals, we compare the empirical likelihood and the normal approximation methods with the optimal bandwidth selected by cross-validation. The empirical likelihood method has a better performance most of the time. Finally, a real clinical trial data is used to illustrate how to generate empirical likelihood confidence bands using bootstrap method.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In probability theory, a useful quantity related to cumulative distribution function F is the quantile function. Assuming a continuous and strictly monotonic distribution function, $F: R \to (0,1)$, there is an unique number x_p for each p such that $F(x_p) = P(X \le x_p) = p$. x_p is a number on the set of values of the random variable X, such that the fraction p of the total probability is assigned to the interval $(-\infty, x_p)$. x_p is called a pth quantile of the distribution. In general, the pth quantile of F is defined by $F^{-1}(p) = \inf\{x \in R : p \le F(x)\}$. The analysis of $F^{-1}(p)$ has played one of the major roles in statistics, economics, finance, risk management, etc.

Empirical likelihood (EL) is a nonparametric method of statistical inference based on a data-driven likelihood ratio function. The empirical likelihood method is one of the most famous methodologies for nonparametric inference. The deployment of the EL method with respect to survival analysis can be traced back to Thomas and Grunkemeier (1975). The EL method was completely summarized in Owen (1988, 1990). In addition, EL confidence bands for quantile function and survival function have been derived in Li et al. (1996) and in Hollander et al. (1997), respectively. For complete data, Chen and Hall (1993) developed the smoothed EL inference for quantiles. Jing (1995), Qin and Zhao (1997) and Qin (1997) respectively proposed methodologies for the confidence intervals of the difference between two sample means and quantiles. For censored data, McKeague and Zhao (2005) proposed EL simultaneous confidence band for the difference of distribution functions in two-sample case. Zhao and Zhao (2011) developed EL confidence intervals for the contrast of two hazard functions motivated by Shen and He (2008). Moreover, Zhou and Jing (2003) proposed the smoothed EL method for the difference of quantiles for one sample complete data. Then Shen and He (2007)

* Corresponding author.
E-mail addresses: matyiz@langate.gsu.edu, yichuan@gsu.edu (Y. Zhao).

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.09.010

proposed a smoothed EL method on the same setting for one sample with right censoring. In this paper, we investigate the interval estimate of the quantile difference of two samples $\theta_0 = F_1^{-1}(p) - F_2^{-1}(p)$. Adopting the method proposed by Shen and He (2007), we develop the EL confidence interval for the difference of quantiles θ_0 with two samples.

The paper is organized as follows. In Section 2, the empirical likelihood procedure for the θ_0 is developed. In Section 3, simulation studies are presented for comparing the performance between the smoothed empirical likelihood confidence intervals and the traditional normal approximation confidence intervals. It is then followed by an analysis on a real data application in Section 4. Section 5 is a summary of conclusions and discussion. The proofs are contained in the Appendix.

2. Main results

Throughout this paper, we are using the standard two samples with right censored data. We adopt the same notations as Shen and He (2007) and Yang and Zhao (2012) did for simplicity. First of all, let $T_{ji} \geq 0$, $i = 1, \ldots, n_j$ be i.i.d. failure times with continuous distribution F_j, $j=1,2$. Let C_{ji}, $i = 1, \ldots, n_j$ be i.i.d. censoring times with continuous distribution function G_j, $j=1,2$, and independent of T_{ji}. We observe each sample in the form of (X_{ji}, δ_{ji}), where $X_{ji} = \min(T_{ji}, C_{ji})$ and $\delta_{ji} = I(T_{ji} \leq C_{ji})$. Like Shen and He (2007), we denote $X_{(ji)}$ as the order statistics of the jth sample. The concomitant of δ_{ji} is denoted as $\delta_{(ji)}$ and the number of risk sets is r_{ji}. Let $\eta_0 = F_2^{-1}(p)$. Since F is strictly monotone increase function and $\theta_0 = \theta(p) = F_1^{-1}(p) - F_2^{-1}(p)$, we have $F_1^{-1}(p) = \eta_0 + \theta_0, F_2^{-1}(p) = \eta_0$. The likelihood ratio is defined as Shen and He (2007) did similarly,

$$R(\theta_0, \eta, p) = \frac{\sup_{\varphi_{ji} \in \Phi}\{L(F_1, F_2) : F_1(\eta + \theta_0) = p, F_2(\eta) = p\}}{\sup_{\varphi_{ji} \in \Phi} L(F_1, F_2),}$$

where $\varphi_{j1}, \varphi_{j2}, \ldots, \varphi_{jn_j}$ are the hazard values at $X_{(j1)}, X_{(j2)}, \ldots, X_{(jn_j)}$ and $\varphi_j = (\varphi_{j1}, \varphi_{j2}, \ldots, \varphi_{jn_j}) \in \Phi$, the space of hazard values. It is computationally difficult to maximize $R(\theta_0, \eta, p)$ and a smoothed $\tilde{R}(\theta_0, \eta, p)$ is proposed in this case. Let $K(t)$ be a smooth distribution function, and h_j be a bandwidth. Denote $K_{h_j}(t) = K(t/h_j)$. Like Shen and He (2007) and Yang and Zhao (2012), we define

$$\Phi_2 = \left\{ \varphi_1, \varphi_2 \in \Phi : \sum_{i=1}^{n_1} K_{h_1}(\eta + \theta_0 - X_{(1i)}) \ln(1 - \varphi_{1i}) = \ln(1-p), \sum_{i=1}^{n_2} K_{h_2}(\eta - X_{(2i)}) \ln(1 - \varphi_{2i}) = \ln(1-p) \right\}.$$

And hence the smoothed EL ratio $\tilde{R}(\theta_0, \eta, p)$ is given like Shen and He (2007) and Yang and Zhao (2012). Define $\hat{R}(\theta_0, p) = \sup_\eta \tilde{R}(\theta_0, \eta, p)$. Denote

$$q_{1n_1}(\eta, \lambda_1) = \sum_{i=1}^{n_1} K_{h_1}(\eta + \theta_0 - X_{(1i)}) \ln\left(1 - \frac{\delta_{(1i)}}{r_{1i} + \lambda_1 K_{h_1}(\eta + \theta_0 - X_{(1i)})}\right) - \ln(1-p) = 0, \tag{2.1}$$

$$q_{2n_2}(\eta, \lambda_2) = \sum_{i=1}^{n_2} K_{h_2}(\eta - X_{(2i)}) \ln\left(1 - \frac{\delta_{(2i)}}{r_{2i} + \lambda_2 K_{h_2}(\eta - X_{(2i)})}\right) - \ln(1-p) = 0, \tag{2.2}$$

$$\begin{aligned} q_{3n_1n_2}(\eta, \lambda_1, \lambda_2) = &\sum_{i=1}^{n_1} \lambda_1 K'_{h_1}(\eta + \theta_0 - X_{(1i)}) \ln\left(1 - \frac{\delta_{(1i)}}{r_{1i} + \lambda_1 K_{h_1}(\eta + \theta_0 - X_{(1i)})}\right) \\ &+ \sum_{i=1}^{n_2} \lambda_2 K'_{h_2}(\eta - X_{(2i)}) \ln\left(1 - \frac{\delta_{(2i)}}{r_{2i} + \lambda_2 K_{h_2}(\eta - X_{(2i)})}\right) = 0, \end{aligned} \tag{2.3}$$

$$\begin{aligned} \ln \hat{R}(\theta_0, p) = &\sum_{i=1}^{n_1} (r_{1i} - \delta_{(1i)}) \ln\left(1 + \frac{\lambda_1 K_{h_1}(\eta + \theta_0 - X_{(1i)})}{r_{1i} - \delta_{(1i)}}\right) - r_{1i} \ln\left(1 + \frac{\lambda_1 K_{h_1}(\eta + \theta_0 - X_{(1i)})}{r_{1i}}\right) \\ &+ \sum_{i=1}^{n_2} (r_{2i} - \delta_{(2i)}) \ln\left(1 + \frac{\lambda_2 K_{h_2}(\eta - X_{(2i)})}{r_{2i} - \delta_{(2i)}}\right) - r_{2i} \ln\left(1 + \frac{\lambda_2 K_{h_2}(\eta - X_{(1i)})}{r_{2i}}\right), \end{aligned} \tag{2.4}$$

where η, λ_1 and λ_2 satisfy Eqs. (2.1)–(2.3). Now we have the following theorems.

Theorem 2.1 (*Yang and Zhao, 2012*). *Assume regularity conditions* (C.1)–(C.5) *in the Appendix hold. Then, the maximum of* $L(F_1, F_2)$ *with the constraint condition* Φ_2 *for large* $n = n_1 + n_2$ *is achieved at a unique* φ_1, φ_2, *a.s.*

Theorem 2.2. *Assume regularity conditions* (C.1–C.5) *in the Appendix hold. For fixed* p, *it exists a solution* η_E, *i.e.,* $\tilde{R}(\theta_0, \eta_E, p) = \sup_\eta \tilde{R}(\theta_0, \eta, p) = \hat{R}(\theta_0, p)$. *We have*

$$-2 \ln \tilde{R}(\theta_0, \eta_E, p) \xrightarrow{\mathfrak{D}} \chi_1^2.$$

Download English Version:

https://daneshyari.com/en/article/1149148

Download Persian Version:

https://daneshyari.com/article/1149148

Daneshyari.com