Available online at www.sciencedirect.com

ScienceDirect

Statistics & Probability Letters 77 (2007) 549–557

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Empirical likelihood inference for the mean residual life under random censorship

Gengsheng Qin*, Yichuan Zhao

Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA

Received 6 January 2005; received in revised form 1 September 2006; accepted 20 September 2006
Available online 4 December 2006

Abstract

Mean residual life (MRL) for a lifetime random variable X is one of the basic parameters of interest in survival analysis. There has been a lot of work available on the inference of MRL in the complete data setting. However, the observations for X are often censored. Inference for MRL becomes more involved under random censorship. In this paper, an empirical likelihood procedure is proposed for the inference of MRL with right censored data. It is shown that the limiting distribution of the empirical log-likelihood ratio for MRL is a scaled chi-square distribution. The limiting distribution can be used to construct empirical likelihood-based confidence intervals for MRL. Numerical results from a simulation study suggest that the empirical likelihood-based confidence intervals have better coverage accuracy than the existing normal approximation-based confidence intervals.

Keywords: Empirical likelihood; Censored data; Mean residual life; Kaplan–Meier estimator

1. Introduction

One of the basic parameters of interest in survival analysis is the *mean residual life* (MRL) at time x. For individuals of age x, this parameter measures their expected remaining lifetime. Let X be a continuous lifetime variable with cdf F. The corresponding MRL at time x is defined as

$$M(x) \equiv E[X - x|X \geqslant x] = \int_x^\tau S(t)\,\mathrm{d}t/S(x),$$

where $S(x) = 1 - F(x)$ is the survival rate at time x, $\tau = \inf\{x : S(x) = 0\}$.

Comprehensive reviews of MRL function have been made available by Proschan and Serfling (1974) and Guess and Proschan (1988). Because of the closed relationship between $M(x)$ and $S(x)$, non-parametric function estimation methods can lend themselves to construct an estimator for $M(x)$ under the complete data setting (Yang, 1978; Hall and Wellner, 1979; Csörgö and Zitikis, 1996).

*Corresponding author.
E-mail address: gqin@gsu.edu (G. Qin).

0167-7152/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2006.09.018

Censored data often arise in the study of medical follow-up, survival analysis and reliability study. Let $X_1, \ldots, X_n$ be an iid copy of random variable X, and let $Y_1, \ldots, Y_n$ be an iid copy of censoring time Y with cdf G. Assume X_i's and Y_i's are independent. The observations in the random censoring model are

$$Z_i = \min(X_i, Y_i), \quad \delta_i = I(X_i \leqslant Y_i), \quad i = 1, \ldots, n.$$

Based on the censored data (Z_i, δ_i)'s, inference for $M(x)$ becomes more involved. Yang (1977) proposed the following estimator for $M(x)$:

$$\widehat{M}_0(x) = \int_x^T \widehat{S}(t)\,\mathrm{d}t / \widehat{S}(x),$$

where T is a constant with $F(T) < 1$, $\widehat{S}(t) = 1 - \widehat{F}(t)$, and $\widehat{F}$ is the Kaplan–Meier estimator of F, i.e.,

$$\widehat{F}(t) = 1 - \prod_{s \leqslant t} (1 - \mathrm{d}N(s)/Y(s))$$

with $N(s) = \sum_{i=1}^n I(Z_i \leqslant s, \delta_i = 1)$, and $Y(s) = \sum_{i=1}^n I(Z_i \geqslant s)$. Kumazawa (1987) proposed the estimator with T replaced by $Z_{(n)} = \max(Z_1, \ldots, Z_n)$,

$$\widehat{M}(x) = \int_x^{Z_{(n)}} \widehat{S}(t)\,\mathrm{d}t / \widehat{S}(x), \tag{1}$$

and proved that the distribution of $\sqrt{n}(\widehat{M}(x) - M(x))$ at time x is asymptotically normal with mean 0 and variance $V^2(x)$. He also gave an estimator for the variance,

$$\widehat{V}^2(x) = \frac{n}{\widehat{S}^2(x)} \int_x^{Z_{(n)}} \left(\int_s^{Z_{(n)}} \widehat{S}(t)\,\mathrm{d}t \right)^2 \frac{I(Y(s) > 0)\,\mathrm{d}N(s)}{Y(s)(Y(s) - 1)}. \tag{2}$$

Therefore, confidence interval for the MRL at time x can be constructed by using the normal approximation (NA) method. Although the NA-based interval is useful in some situations, there are two main drawbacks associated with the variance estimator. First, $\widehat{V}^2(x)$ takes a complicated form and may not be stable, particularly when x is near the tail of lifetime distribution. For an unstable variance estimate, it is possible that the confidence interval for the MRL will contain negative values. Second, the method does not always perform well for small samples. Our simulation study in this paper showed that this variance estimator was too conservative, which led to the heavy overcoverage of the NA-based confidence intervals.

Empirical likelihood (EL) was introduced by Owen (1988). EL approach is more accurate than the NA in many situations particularly when the underlying distribution is non-normal and the variance estimate is unstable. Its advantages have been well recognized (Hall and La Scala, 1990). EL method has numerous applications in various areas of statistics (Chen, 1993, 1996; Chen and Hall, 1993; Qin and Lawless, 1994). Recently, it has been applied in various survival analysis contexts (Adimari, 1997; Qin and Jing, 2001a–c; McKeague and Zhao, 2002; Qin and Tsao, 2003).

In this paper, we develop an EL approach for the inference of the MRL with right censored data. We first propose an estimating equation for the MRL, and then define an EL ratio for the MRL and show that its limiting distribution is a scaled chi-square distribution with one degree of freedom. This limiting distribution is used to construct an EL-based confidence interval for the MRL. Our simulation results suggest that the EL-based interval is more accurate than the NA-based interval in terms of coverage probability.

The present paper is organized as follows. The EL theory for the MRL is presented in Section 2. In Section 3, a simulation study is conducted to compare the proposed EL method with the existing NA method. The proof of the main result is deferred until the Appendix.

2. Methodology and main result

Note that the MRL at time x can be rewritten as

$$M(x) = E[XI(X \geqslant x)]/E[I(X \geqslant x)] - x.$$