Statistics and Probability Letters 83 (2013) 2332–2336

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A note on estimation of the mean residual life function with left-truncated and right-censored data

Mu Zhao [a,b], Hongmei Jiang [a,*], Xu Liu [a]

[a] Department of Statistics, Northwestern University, Evanston 60208, United States
[b] School of Mathematical Sciences, Anhui University, Hefei 230039, China

ARTICLE INFO

Article history:
Received 11 January 2013
Received in revised form 16 June 2013
Accepted 16 June 2013
Available online 24 June 2013

Keywords:
Mean residual life function
Left-truncated and right-censored data
TJW product-limit estimator
Weak convergence

ABSTRACT

This note focuses on estimating the mean residual life function with left-truncated and right-censored data. We show that the proposed estimator converges weakly to a Gaussian process. The performances of the estimator and its pointwise confidence intervals are illustrated through simulation studies.

1. Introduction

The mean residual life (MRL) function of a non-negative random variable T at time $t \geqslant 0$ is defined as

$$m(t) = E(T - t|T > t) = \frac{\int_t^\infty S(u)du}{S(t)} I(S(t) > 0), \tag{1.1}$$

where $S(t)$ is the survival function of T. This function is of interest in many fields such as reliability research, survival analysis, actuarial study, and so forth. For instance, an insurance company may be concerned about how much longer a product can be used, given that the product has been used normally for, say, t years. Estimation of $m(t)$ was first considered by Yang (1978), replacing the survival function in (1.1) with its empirical survival function. The properties of this estimator were discussed by Csörgo and Zitikis (1996). A natural extension of this estimator to randomly right-censored data is replacing $S(t)$ in (1.1) with its Kaplan–Meier estimator; see, for example, Yang (1977) and Hall and Wellner (1981) and references therein. Chaubey and Sen (1999, 2008) proposed an alternative smoothed estimator based on complete data and right-censored data respectively.

However, in epidemiology and individual follow-up studies, some individuals may be subject to left truncation. Left truncation occurs when a subject is not included in the study if the corresponding failure event precedes the starting time of the studies. Here, we use Channing House data (Hyde, 1980) as an illustrative example. Channing House is a retirement center in Palo Alto, California. 97 men and 365 women passed through the center from the opening of the house in 1964 until July 1, 1975, when the data were collected for the purpose of studying differences in survival of the sexes. Individuals aged 60 years or older and covered by the health program of the center were eligible to enter the retirement community. This implies that the interest in this example is in studying the survival time of those residents who were born before July 1,

* Corresponding author. Tel.: +1 847 467 1087; fax: +1 847 491 4939.
E-mail address: hongmei@northwestern.edu (H. Jiang).

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.06.020

1915, and covered by the health program. Let T be the time from birth to death and V be the (potential) entry age. Then, only individuals with $T \geqslant V$ can be part of the sample, which results in left truncation. In this example, individuals who are subject to left truncation were further subject to right censoring due to the termination of the follow-up. Other examples of left-truncated data can also be found in cases such as an AIDS blood-transfusion study (Wang, 1989), Canadian dementia studies (Asgharian and Wolfson, 2005; Shen et al., 2009), and an electrical transmission study (Hong et al., 2009).

The aim of this paper is to estimate the MRL function with left-truncated and right-censored (LTRC) data. This article is organized as follows. We present some notation and assumptions in Section 2. The main results are stated in Section 3. The performance of the proposed estimator and the comparisons between an NA-based confidence interval and a bootstrap confidence interval are reported in Section 4. Some concluding remarks are presented in Section 5.

2. Notation and assumptions

Let (T, V, C) denote a random vector, where T is the random lifetime variable of interest with continuous survival function $S(t) = 1 - F(t)$, V is the left truncation random variable with continuous d.f. G, and C is the right censoring random variable with continuous d.f. L. It is assumed that T, V and C are mutually independent. One observes (Y, V, δ) if $Y \geqslant V$, where $Y = T \wedge C = \min(T, C)$ and $\delta = I(T \leqslant C)$. When $Y < V$, nothing is observed. Let $\alpha = P(Y \geqslant V) > 0$ and W denote the d.f. of Y. Suppose $\{(Y_i, V_i, \delta_i), 1 \leqslant i \leqslant n\}$ to be a sequence of independent and identically distributed (i.i.d.) random samples of (Y, V, δ) which one observes (i.e. $Y_i \geqslant V_i$). For any d.f. function K, let a_K and b_K denote the left and right endpoints of its support. For identification of F, we need to assume that $a_G \leqslant a_W$, $b_G \leqslant b_W$ (see also Gibels and Wang (1993) and Zhou and Yip (1999)). Meanwhile, for brevity, throughout this article, we assume that

$$a_G = a_W = 0, \qquad b_G \leqslant b_W. \tag{2.1}$$

Define $M(t) = \int_t^{\infty} S(z)dz$ and $C(z) = P(V \leqslant z \leqslant Y | Y \geqslant V) = \alpha^{-1} P(V \leqslant z \leqslant C)(1 - F(z))$. A consistent estimator of $C(z)$ is given by its empirical analog

$$C_n(z) = n^{-1} \sum_{i=1}^{n} I(V_i \leqslant z \leqslant Y_i).$$

As discussed by Zhou and Yip (1999), for any $0 < b < b_W$, we assume that

$$\int_0^b \frac{dW_1(z)}{C^3(z)} < \infty, \tag{2.2}$$

where $W_1(y) = P(Y \leqslant y, \delta = 1 | V \leqslant Y)$, which can be consistently estimated by

$$W_{n1}(y) = n^{-1} \sum_{i=1}^{n} I(Y_i \leqslant y, \delta_i = 1).$$

3. The estimator $\widehat{m}(t)$ and its weak convergence

A nonparametric maximum likelihood estimator of $S(t)$, based on (Y_i, V_i, δ_i), $i = 1, 2, \ldots, n$, is the TJW product-limit estimator $\widehat{S}_n(t)$ which is defined in Tsai et al. (1987) as

$$\widehat{S}_n(t) = \begin{cases} \prod_{Y_i \leq t} \left[1 - (nC_n(Y_i))^{-1}\right]^{\delta_i}, & \text{if } t < Y_{(n)}, \\ 0, & \text{if } t \geqslant Y_{(n)}, \end{cases} \tag{3.1}$$

where $Y_{(n)} = \max(Y_1, \ldots, Y_n)$. On the basis of $\widehat{S}_n(t)$ in (3.1), $m(t)$ can be estimated by

$$\widehat{m}(t) = \widehat{S}_n^{-1}(t) \int_t^{\infty} \widehat{S}_n(u) du I(t \leqslant Y_{(n)}). \tag{3.2}$$

Next, we shall establish that $\widehat{m}(t)$ convergences weakly to a Gaussian process by using the functional delta method. Lemma 1 shown below gives the asymptotic behavior of TJW product-limit estimator $\widehat{S}_n(t)$. The proof can be found in Zhou and Yip (1999). Lemma 2 presents a functional delta method which can be found in Van der Vaart (2000) and Kosorok (2008).

Lemma 1. *Define $Z_n(t) = n^{1/2}(\widehat{S}_n(t) - S(t))$, $\tau < \infty$ with $S(\tau) > 0$. Then, for $0 < t < \tau$, $Z_n(t)$ converges in distribution to a Gaussian process $Z(t)$ with mean zero and covariance function*

$$\gamma(s, t) = S(s)S(t) \int_0^s \frac{dW_1(z)}{C^2(z)}, \quad s \leqslant t, \tag{3.3}$$

and $Z(t)$ has a continuous sample path with probability 1.