Statistics and Probability Letters 82 (2012) 626–635

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

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

Model-based likelihood ratio confidence intervals for survival functions

Sundarraman Subramanian *

Center for Applied Mathematics and Statistics, Department of Mathematical Sciences, New Jersey Institute of Technology, 07102, United States

ARTICLE INFO

Article history:
Received 5 August 2011
Received in revised form 28 November 2011
Accepted 29 November 2011
Available online 8 December 2011

Keywords:
Gaussian process
Lagrange multiplier
Maximum likelihood estimate
Misspecified model
Semiparametric random censorship model

ABSTRACT

We introduce an adjusted likelihood ratio procedure for computing pointwise confidence intervals for survival functions from censored data. The test statistic, scaled by a ratio of two variance quantities, is shown to converge to a chi-squared distribution with one degree of freedom. The confidence intervals are seen to be a neighborhood of a semiparametric survival function estimator and are shown to have correct empirical coverage. Numerical studies also indicate that the proposed intervals have smaller estimated mean lengths in comparison to the ones that are produced as a neighborhood of the Kaplan–Meier estimator. We illustrate our method using a lung cancer data set.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Suppose that, for an unknown survival function $S(t)$ and a fixed $a > 0$, we are interested in computing a confidence interval for $S(a)$. When data are right censored, a well-known approach is to apply the asymptotic normality of the Kaplan–Meier (KM) estimator together with the Greenwood estimator of its variance. Alternatively, one may utilize a semiparametric random censorship model (SRCM) based estimator of $S(t)$ introduced by Dikta (1998). When the parametric component is specified correctly the latter estimator has a smaller asymptotic variance than the KM estimator, and, hence, would produce a tighter confidence interval for $S(a)$.

A likelihood ratio (LR) confidence interval for $S(a)$, on the other hand, is formed by collecting all values p for which the null hypothesis $H_0 : S(a) = p$ is not rejected by a LR test. For right censored data, Thomas and Grunkemeier (1975) introduced an LR test statistic and obtained an LR confidence interval for $S(a)$. Li (1995a,b) further studied this problem for right censored data and left truncated data respectively, and Murphy (1995) considered other generalizations. The nonparametric LR confidence interval for $S(a)$ may be interpreted as a neighborhood of $\hat{S}_{KM}(a)$, the KM estimator evaluated at the point a. However, in light of the aforementioned efficiency of $\hat{S}(t)$, the SRCM-based estimator, it is desirable to obtain a neighborhood of $\hat{S}(a)$, which would yield a tighter confidence interval than the nonparametric method based on the KM estimator. In this article we propose and implement an SRCM-based adjusted LR test for improved interval estimation of $S(a)$. We show that the proposed method offers significant improvements over the nonparametric method.

The article is organized as follows. In Section 2, we introduce notation, review the nonparametric LR procedure, and then develop our proposed modification. We present simulation results and a real example illustration in Section 3. We present a concluding discussion in Section 4, where we outline an extension to handle missing censoring indicators.

* Tel.: +1 973 642 4496.
E-mail address: sundars@njit.edu.

0167-7152/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2011.11.028

2. Proposed model-based procedure

Let $T_1, \ldots, T_n$ be the i.i.d. failure times, having the common survival function $S(t)$. Let $C_1, \ldots, C_n$ be the i.i.d. censoring times with common distribution function $G(t)$, and independent of the T_i's. We observe $\{(X_i, \delta_i), 1 \le i \le n\}$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \le C_i)$. Let $\Delta W(t) = W(t) - W_-(t)$, where $W_-(t) = W(t-)$. Writing $N^u(t) = \sum_{i=1}^n I(X_i \le t, \delta_i = 1)$ and $Y(t) = \sum_{i=1}^n I(X_i > t)$, the KM estimator of $S(t)$ over $[0, \infty)$ is given by

$$\hat{S}_{\mathrm{KM}}(t) = \prod_{s \le t} \left(1 - \frac{\Delta N^u(s)}{Y(s)}\right), \tag{2.1}$$

where $\Delta N^u(s)/Y(s)$ is defined as 0 when $Y(s) = 0$.

2.1. Review of nonparametric likelihood ratio confidence intervals

Let $\mathbf{\Gamma}$ denote the space of all survival functions on $[0, \infty)$. It is well known (e.g., Li, 1995a) that $\hat{S}_{\mathrm{KM}}$ maximizes the likelihood function $L(S)$ over $\mathbf{\Gamma}$, where

$$L(S) = \prod_{i=1}^{n} [S_-(X_i) - S(X_i)]^{\delta_i} [S(X_i)]^{1-\delta_i}. \tag{2.2}$$

In particular, the maximum value $L(\hat{S}_{\mathrm{KM}})$ is given by

$$L\left(\hat{S}_{\mathrm{KM}}\right) = \prod_{s} \left[\frac{\Delta N^u(s)}{Y(s)}\right]^{\Delta N^u(s)} \left[1 - \frac{\Delta N^u(s)}{Y(s)}\right]^{Y(s) - \Delta N^u(s)}. \tag{2.3}$$

For a nonparametric LR test, we also need to obtain the constrained estimator, denoted by $\hat{S}_{\mathrm{KMC}}$, that maximizes $L(S)$ over $\mathbf{\Gamma}$ subject to $S(a) = p$. This estimator is given by

$$\hat{S}_{\mathrm{KMC}}(t) = \begin{cases} \prod_{s \le t} \left(1 - \dfrac{\Delta N^u(s)}{Y(s) + \hat{\lambda}}\right) & \text{if } t \le a \\ \prod_{s \le a} \left(1 - \dfrac{\Delta N^u(s)}{Y(s) + \hat{\lambda}}\right) \prod_{a < s \le t} \left(1 - \dfrac{\Delta N^u(s)}{Y(s)}\right) & \text{if } t > a, \end{cases}$$

see Thomas and Grunkemeier (1975). Here $\hat{\lambda}$ solves

$$\prod_{s \le a} \left(1 - \frac{\Delta N^u(s)}{Y(s) + \lambda}\right) = p. \tag{2.4}$$

In particular, the maximum value $L(\hat{S}_{\mathrm{KMC}})$ is given by

$$L\left(\hat{S}_{\mathrm{KMC}}\right) = \prod_{s \le a} \left[\frac{\Delta N^u(s)}{Y(s) + \hat{\lambda}}\right]^{\Delta N^u(s)} \left[1 - \frac{\Delta N^u(s)}{Y(s) + \hat{\lambda}}\right]^{Y(s) - \Delta N^u(s)} \prod_{s > a} \left[\frac{\Delta N^u(s)}{Y(s)}\right]^{\Delta N^u(s)} \left[1 - \frac{\Delta N^u(s)}{Y(s)}\right]^{Y(s) - \Delta N^u(s)}. \tag{2.5}$$

From Eqs. (2.3) and (2.5), it is seen that $\log \tilde{R} = \log \left(L(\hat{S}_{\mathrm{KMC}})/L(\hat{S}_{\mathrm{KM}})\right)$ leads to Eq. (3.6) of Thomas and Grunkemeier (1975), expressed in our notation as follows:

$$\log \tilde{R} = \sum_{s \le a} \left[\left(Y(s) - \Delta N^u(s)\right) \log \left(1 + \frac{\hat{\lambda}}{Y(s) - \Delta N^u(s)}\right) - Y(s) \log \left(1 + \frac{\hat{\lambda}}{Y(s)}\right)\right]. \tag{2.6}$$

See also Eq. (14) of Hollander et al. (1997). The quantity $-2 \log \tilde{R}$ converges in distribution to χ_1^2 (e.g., Li, 1995a). From this result, we can collect all λ that, when substituted for $\hat{\lambda}$ in Eq. (2.6), give values that are at most $\chi_1^2(\alpha)$, for a pre-specified $0 < \alpha < 1$. Here, $\chi_1^2(\alpha)$ is the upper α quantile of the chi-squared distribution with 1 degree of freedom. Thomas and Grunkemeier (1975) showed that the collection of such λ is a closed interval $[\lambda_L, \lambda_U]$, with $\lambda_L < 0 < \lambda_U$ when $0 < \hat{S}_{\mathrm{KM}}(a) < 1$ and $0 = \lambda_L < \lambda_U$ when $\hat{S}_{\mathrm{KM}}(a) = 0$. The nonzero limits λ_L and λ_U are obtained by solving $-2 \log \tilde{R} = \chi_1^2(\alpha)$. Because p is an increasing function of λ, see Eq. (2.4), the confidence interval for p is a closed interval $[p_L, p_U]$, where $0 < p_L < p_U < 1$ when $0 < \hat{S}_{\mathrm{KM}}(a) < 1$, and $0 = p_L < p_U < 1$ when $\hat{S}_{\mathrm{KM}}(a) = 0$, and p_L and p_U are given by

$$p_L = \prod_{s \le a} \left(1 - \frac{\Delta N^u(s)}{Y(s) + \lambda_L}\right), \qquad p_U = \prod_{s \le a} \left(1 - \frac{\Delta N^u(s)}{Y(s) + \lambda_U}\right).$$

Download English Version:

https://daneshyari.com/en/article/1154906

Download Persian Version:

https://daneshyari.com/article/1154906

Daneshyari.com