



Parameter estimation of partial linear model under monotonicity constraints with censored data

Wei Chen^{a,b,*}, Xiaojia Li^c, Dehui Wang^a, Guohua Shi^b

^a College of Mathematics, Jilin University, Changchun 130012, PR China

^b School of Zhangjiagang, Jiangsu University of Science and Technology, Zhangjiagang 215600, PR China

^c School of Business, Jilin University, Changchun 130012, PR China

ARTICLE INFO

Article history:

Received 13 June 2014

Accepted 23 December 2014

Available online 13 January 2015

AMS 2000 subject classifications:

62N01

62G08

Keywords:

B-splines

Censored data

Kaplan–Meier weights

Least square

Monotonicity constraints

ABSTRACT

We propose a weighted least square method for estimation in the partial linear model with monotonicity constraints and right-censored data. This method uses the Kaplan–Meier weights to account for censoring and monotone B-splines to approximate the unknown monotone function. We show that the proposed estimator of regression coefficients is root- n consistent and asymptotically normal under appropriate assumptions. One advantage is that our method can be easily computed using existing software. A simulation study is conducted to evaluate the finite sample performance of the proposed method.

© 2014 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.

1. Introduction

In survival analysis, it is often of interest to explore the relationship between the failure time and a collection of covariates. For this purpose, a large number of semiparametric regression models and estimation methods have been developed. Among them, Cox's (1972) proportional hazards (PH) model may be the most popular and widely-used statistical tool for analyzing survival data partly due to the efficient inference based on the partial likelihood and the availability of implementation in almost all existing software. However, the PH model requires that the hazard ratio is always constant over time between any two subjects with distinct covariates. In some situations, this assumption seems to be rather restrictive and hard to be met. See Struthers and Kalbfleisch (1986). Thus, other alternatives to the PH model with nonproportional risks or more flexibility of modeling the covariates are desirable.

As suggested by Sir David Cox, the accelerated failure time (AFT) model is, “in many ways more appealing because of its quite direct physical interpretation”. The AFT model assumes that the logarithm of survival time is linearly correlated to a vector of covariates of interest, which can be specified as

$$\log \tilde{T} = \beta'X + \epsilon,$$

where $\tilde{T}$ denotes the failure time, β is a p -vector of regression coefficients to be estimated, X is a p -vector of covariates, ϵ is the random error with mean zero but unknown distribution, and the superscript ($'$) denotes the transpose of a column

* Corresponding author at: School of Zhangjiagang, Jiangsu University of Science and Technology, Zhangjiagang 215600, PR China. Tel.: +86 13578653772; fax: +86 13578653772.

E-mail address: novicejlu@gmail.com (W. Chen).

<http://dx.doi.org/10.1016/j.jkss.2014.12.002>

1226-3192/© 2014 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.

vector. In the presence of right censored data, several semiparametric estimates have been proposed, such as the least square estimator (Buckley & James, 1979; Ritov, 1990) and the rank-based estimator (Fygenson & Ritov, 1994; Tsiatis, 1990; Ying, 1993). Especially, for the rank-based estimator, Jin, Lin, and Wei (2003) provided the first reliable and accurate computing procedure via linear programming (LP) to obtain the Gehan estimator.

Although the AFT model is useful, the assumption that each covariate has a linear effect on the log survival time is not appropriate in some situations. For example, in many clinical trials and biomedical studies, one is primarily concerned about identifying the effect of a treatment when a confounding factor of less interest exists. In such cases, it is reasonable and useful to treat the confounding factor as a nonparametric component without loss of the easy interpretation of the treatment effect (Chen, Shen, & Ying, 2005). In the literature, one usually characterizes these covariate effects through a model referred to as the partial linear (PL) model, which can be written as

$$\log \tilde{T} = \beta'X + g(Z) + \epsilon, \quad (1)$$

where Z is a univariate covariate such as the confounding factor, $g(\cdot)$ is an unknown smooth function playing a role of the nonparametric component, and other notations are defined as above. When the response variable $\tilde{T}$ is completely observed, many researchers have studied the PL model, see Hardle and Liang (2007). When the response variable is subject to right censoring, several authors have investigated the inference of model (1). Orbe, Ferreira, and Nez-Anton (2003) adapted Stute's (1993) method and proposed a penalized weighted least square method with the unknown function $g(\cdot)$ being approximated by the cubic splines. Chen et al. (2005) developed a strategy to eliminate the function $g(\cdot)$ by a proper stratification, thus proposed an estimation method, which is a Gehan-type extension of the Wilcoxon–Mann–Whitney estimating function. Zou, Zhang, and Qin (2011) incorporated the penalized splines into the Gehan-type estimating function that occurred in the rank-based inference for the AFT model and obtained the estimate of regression coefficients and nonparametric component simultaneously. Recently, Chen, Wang, and Li (2013) proposed a computationally efficient procedure to solve the numerical problem that has arisen in Zou et al. (2011), which utilizes the polynomial-based smoothing method.

In many studies, there exists a monotonic relationship between one covariate and the response variable, i.e., the unknown smooth function $g(\cdot)$ is of monotonicity property. For example, the growth curves and dose–response curves are well-known to be monotone. When the available data are complete, many researchers have investigated the model (1) with monotonicity constraints. Huang (2002) considered the isotonic regression approach and showed his constrained least square estimator of β is root- n consistent and asymptotically normal. Lu (2010) proposed a sieve maximum likelihood estimation based on monotone B-Splines, where he assumed the error term is normally distributed. Du, Sun, and Xie (2013) studied the M-estimation for the model (1) under monotonic constraints. Recently, Yu (2014) considered the partial linear additive isotonic regression model. To the best of our knowledge, however, there is little study for the model (1) with $g(\cdot)$ being monotone and the response variable subject to right censoring. Therefore, it would be preferable to develop a practical method for this case. In this paper, we proposed an easily-implemented estimation method. Specifically speaking, we proposed the spline-based weighted least square estimator of (β, g) when g is subject to be monotone. Specially, the weights used here are the Kaplan–Meier weights introduced initially by Stute (1993) and have been applied generally into the variable selection and estimation issues of the AFT model and the censored quantile regression model (Huang, Ma, and Xie (2006, 2007); Hu & Chai, 2013, among others). Moreover, estimation of an unknown monotone function via splines is also common in the literature of survival analysis. For example, Shen (1998) developed a splines-based sieve maximum likelihood estimation method for the baseline cumulative hazard function and the regression coefficients in the proportional odds (PO) model with right censored data and case-2 interval censored data. Lu, Zhang, and Huang (2007) considered a nonparametric monotone I-spline method for the proportional mean model with panel count data. In the presence of case-1 interval censored data, McMahan, Wang, and Tebbbs (2013) used monotone splines to model the baseline cumulative hazard function in the PH model and the baseline odds function in the PO model, and developed a computationally feasible algorithm.

The rest of this paper is organized as follows. In Section 2, we introduce the Kaplan–Meier weights and the B-spline representation of the unspecified monotone function, and derive a weighted least square loss function and provide a computationally feasible algorithm to solve the constrained optimization problem. Under some regularity conditions, the resulting estimator of regression parameters is shown to be root- n consistent and asymptotically normal in Section 3. The proposed procedure is implemented in our simulation studies presented in Section 4. The results show that our method performs quite well, no matter for the estimation of the regression coefficients or the unknown function $g(\cdot)$. Finally, some discussions in Section 5 conclude the paper.

2. Estimation and algorithm

Let T and C denote the logarithm of the failure time and the censoring time, respectively. Let (T_i, C_i, X_i, Z_i) , $i = 1, \dots, n$, be a random sample of (T, C, X, Z) satisfying the model, explicitly,

$$T = \beta'X + g(Z) + \epsilon,$$

where the variables have the same mean as those in (1). Since the failure time is subject to right censoring, we will only observe the vector (Y, δ, X, Z) with $Y = \min(T, C)$ and $\delta = I(T \leq C)$, where $I(\cdot)$ denotes the indicator function. Thus the available data are of the form $(Y_i, \delta_i, X_i, Z_i)$, $i = 1, \dots, n$. As in most cases, we assume (X, Z) and ϵ are independent, and ϵ is zero-mean with finite variance.

Download English Version:

<https://daneshyari.com/en/article/1144656>

Download Persian Version:

<https://daneshyari.com/article/1144656>

[Daneshyari.com](https://daneshyari.com)