Communications for Statistical Applications and Methods
2016, Vol. 23, No. 6, 467–478

http://dx.doi.org/10.5351/CSAM.2016.23.6.467
Print ISSN 2287-7843 / Online ISSN 2383-4757

Semiparametric accelerated failure time model for the analysis of right censored data

Zhezhen Jin[1,a]

[a]Department of Biostatistics, Columbia University, USA

Abstract

The accelerated failure time model or accelerated life model relates the logarithm of the failure time linearly to the covariates. The parameters in the model provides a direct interpretation. In this paper, we review some newly developed practically useful estimation and inference methods for the model in the analysis of right censored data.

Keywords: accelerated failure time model, Buckley-James estimator, censored data, estimating function, Gehan-type estimator, induced smoothing, weighted logrank-type estimator

1. Introduction

The survival data are common in many fields, such as economics, business, industrial engineering and biomedical studies. The nonparametric and semiparametric modeling has been studied extensively because it offers valid estimation and inference with less stringent model assumptions. The Cox proportional hazards model is the most used semiparametric regression models for the analysis of survival data (Cox, 1972) due to the availability of statistical software packages for implementation of the estimation and inference procedures. The accelerated failure time (AFT) models relates the logarithm of the failure time linearly to the covariates. The parameters in the model provides a direct interpretation. However, its estimation and inference procedures are challenging in the presence of censroing. Over the last forty years, there has been considerable research on the AFT models, see Buckely and James (1979), Jin *et al.* (2003, 2006), Koul *et al.* (1981), Lai and Ying (1991a, 1991b, 1995), Miller and Halpern (1982), Powell (1984), Prentice (1978), Ritov (1990), Robins and Tsiatis (1992), Tsiatis (1990), Wei *et al.* (1990), Yang (1997), Ying (1993), Zhou (2005a), Zeng and Lin (2007) and among many others. An earlier review on AFT model can be found in Wei (1992). In this paper, our goal is to review and discuss some newly developed practically useful statistical methods which specifically deal with the semiparametric accelerated failure time model.

2. Semiparametric accelerated failure time model

Let T_i be the failure time for the i^{th} patient, $i = 1, \ldots, n$. Due to censoring C_i, we observe $Y_i = \min(T_i, C_i)$ and $\delta_i = I\{T_i \le C_i\}$ which takes value 1 if $T_i \le C_i$ and 0 otherwise. The semi-parametric AFT model is of form

$$\log(T_i) = X_i^T \beta + \epsilon_i, \tag{2.1}$$

[1] Department of Biostatistics, Columbia University, 722 W 168th Street, New York, NY 10032, USA.
E-mail: zj7@cumc.columbia.edu

Published 30 November 2016 / journal homepage: http://csam.or.kr

where $X_i, i = 1, \ldots, n$ are observed covariates, β is $p \times 1$ unknown parameter vector and $\epsilon_i, i = 1, \ldots, n$ are independent and unobserved errors (the mean of the ϵ_i is not necessarily 0). The semi-parametric AFT model is essentially a regular semiparametric linear regression model and its regression coefficient β is directly quantifies the impact of covariates X on the survival time instead of the more abstract hazard rates. Note that the logarithm in the model may be replaced by any other known strictly increasing functions. However, in the presence of censoring, the intercept parameter is not included in the model specification because the intercept parameter cannot be estimated well as pointed out in Wei (1992).

It is assumed that conditional on the $p \times 1$ covariate vector X_i for the i^{th} subject, the censoring variable C_i and the failure times T_i, $i = 1, \ldots, n$ are independent, and ϵ_i, $i = 1, \ldots, n$ are independent and identically distributed random variables whose common distribution function $F(\cdot)$ is completely unspecified. (Cox and Oakes, 1984; Kalbfleisch and Prentice, 2002; Lawless, 2003).

3. Point estimation

Let $e_i(\beta_0) = \log Y_i - X_i^T\beta_0 = \min\{\log T_i - X_i^T\beta_0, \log C_i - X_i^T\beta_0\} = \min\{\epsilon_i, \log C_i - X_i^T\beta_0\}$, and $N_i(\beta; t) = \Delta_i I\{e_i(\beta) \leq t\}$, where $I(\cdot)$ is the indicator function that takes value 1 when the condition is satisfied 0 otherwise.

3.1. Gehan-type rank estimator

Currently, all available estimation method starts with following estimation equation, based on the Gehan's statistic:

$$U_n^G(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \left(X_i - X_j\right) I\left\{Y_i - Y_j - \left(X_i - X_j\right)^T \beta \leq 0\right\}. \tag{3.1}$$

The function U_n^G is component-wise monotone in β (Fygenson and Ritov, 1994). It was shown that solving $U_n(\beta) = 0$ or $\|U_n(\beta)\| = 0$, is equivalent to minimizing $G_n(\beta)$ (Jin *et al.*, 2003), where

$$G_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \left[Y_i - Y_j - \left(X_i - X_j\right)^T \beta\right]^-, \tag{3.2}$$

where the notation a^- means the negative part of a (e.g. $3^- = 0$, $(-2)^- = 2$). The minimization of $G_n(\beta)$ can be done by following linear programming approach: Minimize $\sum_{i=1}^n \sum_{j=1}^n \Delta_i u_{ij}$ under linear constraints

$$Y_i - Y_j = \left(X_i - X_j\right)^T \beta - u_{ij}, \quad u_{ij} \geq 0.$$

Or alternatively, the minimization of $G_n(\beta)$ is equivalent to minimization of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \left|e_i(\beta) - e_j(\beta)\right| + \left|M - \beta^T \sum_{k=1}^{n} \sum_{l=1}^{n} \delta_k (X_l - X_k)\right|, \tag{3.3}$$

where M is the prespecified extremely large number (Jin *et al.*, 2003). The minimization can be done with many available statistical softwares, such as the function `l1fit` in R.

Therefore, point estimator of β can be obtained, which is denoted as $\hat\beta_G$. Under regularity conditions, it was shown that $\hat\beta_G$ is consistent and $\sqrt{n}(\hat\beta_G - \beta) \sim N(0, \Gamma)$ as $n \to \infty$, where $\Gamma = A_G^{-1} B_G A_G^{-1}$, the A_G and B_G are defined in the expression (3.6) and (3.7) with $\phi(\beta, t) = \sum_{j=1}^n I\{e_j(\beta) \geq t\}$.

3.2. Weighted logrank-type rank estimator

The general weighted log-rank estimating function for β takes the form

$$U_\phi(\beta) = \sum_{i=1}^{n} \Delta_i \phi(\beta; e_i(\beta)) \left\{ X_i - \frac{\sum_{j=1}^{n} X_j I\left\{e_j(\beta) \geq e_i(\beta)\right\}}{\sum_{j=1}^{n} I\left\{e_j(\beta) \geq e_i(\beta)\right\}} \right\} \tag{3.4}$$

or

$$U_\phi(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta; t) \left\{ X_i - \frac{\sum_{j=1}^{n} X_j I\left\{e_j(\beta) \geq t\right\}}{\sum_{j=1}^{n} I\left\{e_j(\beta) \geq t\right\}} \right\} dN_i(\beta; t), \tag{3.5}$$

where ϕ is a possibly data-dependent weight function. When $\phi = 1$, the resulting $U_\phi(\beta)$ corresponds to the log-rank (Mantel, 1966) statistics. When $\phi(\beta, t) = \sum_{j=1}^{n} I\{e_j(\beta) \geq t\}$, the resulting $U_\phi(\beta)$ correspond to the $U_n^G(\beta)$ in (3.1), which is Gehan (1965) statistics.

Under regularity conditions, the root $\hat{\beta}_\phi$ of the estimating function $U_\phi(\beta)$ satisfies that $n^{1/2}(\hat{\beta}_\phi - \beta)$ is asymptotically zero-mean normal with covariance matrix $A_\phi^{-1} B_\phi A_\phi^{-1}$, where

$$A_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi(\beta; t) \left\{ X_i - \frac{\sum_{j=1}^{n} X_j I\left\{e_j(\beta) \geq t\right\}}{\sum_{j=1}^{n} I\left\{e_j(\beta) \geq t\right\}} \right\}^{\otimes 2} \frac{\dot{\lambda}(t)}{\lambda(t)} dN_i(\beta; t), \tag{3.6}$$

$$B_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi^2(\beta; t) \left\{ X_i - \frac{\sum_{j=1}^{n} X_j I\left\{e_j(\beta) \geq t\right\}}{\sum_{j=1}^{n} I\left\{e_j(\beta) \geq t\right\}} \right\}^{\otimes 2} dN_i(\beta; t), \tag{3.7}$$

$A^{\otimes 2} = AA^T$ for a matrix A (Ying *et al.*, 1992), $\lambda(\cdot)$ is the common hazard function of the error terms, and $\dot{\lambda}(t) = d\lambda(t)/dt$ (Lai and Ying, 1991b; Tsiatis, 1990; Wei *et al.*, 1990; Wei, 1992; Ying, 1993).

In general, it is difficult to solve the equation $U_\phi(\beta) = 0$ because $U_\phi(\beta)$ is neither continuous nor componentwise monotone in β. Jin *et al.* (2003) overcame the difficulty by a class of monotone estimation functions that approximate the general weighted log-rank estimating functions (3.4). The idea is to use the Gehan-type rank estimator $\hat{\beta}_G$ in the Section 3.1 and an iterative algorithm. Specifically, Jin *et al.* (2003) proposed to use following modified estimating function of (3.4):

$$\tilde{U}_\phi\left(\beta; \hat{\beta}\right) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \psi\left(\hat{\beta}; t + \left(\beta - \hat{\beta}\right)' X_i\right) S^{(0)}(\beta; t) \left\{ X_i - \frac{\sum_{j=1}^{n} X_j I\left\{e_j(\beta) \geq t\right\}}{\sum_{j=1}^{n} I\left\{e_j(\beta) \geq t\right\}} \right\} dN_i(\beta; t), \tag{3.8}$$

where $\psi(b; x) = \phi(b; x)/S^{(0)}(b; x)$, $S^{(0)}(\beta; t) = \sum_{i=1}^{n} I\{e_i(\beta) \geq t\}$ and $\hat{\beta}$ is a intial consistent estimator of β_0, say $\hat{\beta}_G$. It is easy to see that

$$\tilde{U}_\phi\left(\beta; \hat{\beta}\right) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi\left(\hat{\beta}; e_i\left(\hat{\beta}\right)\right) \Delta_i \left(X_i - X_j\right) I\left\{e_i(\beta) \leq e_j(\beta)\right\}. \tag{3.9}$$

The $\tilde{U}_\phi(\beta; \hat{\beta})$ is monotone in each component of β and is the gradient of the following function:

$$L_\phi\left(\beta; \hat{\beta}\right) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi\left(\hat{\beta}; e_i\left(\hat{\beta}\right)\right) \Delta_i \left\{e_i(\beta) - e_j(\beta)\right\}^{-}. \tag{3.10}$$

Similar to (3.2), the minimisation of $L_\phi(\cdot\,;\hat{\beta})$ can be implemented via linear programming.

The iterative algorithm for estimating β starts from setting: $\hat{\beta}_{(0)} = \hat{\beta}_G$, and $\hat{\beta}_{(k)} = \arg\min_\beta L_\phi(\beta; \hat{\beta}_{(k-1)})$ $(k \geq 1)$. If $\hat{\beta}_{(k)}$ converges to a limit as the number of iterations $k \to \infty$, then the limit must satisfy $U_\phi(\beta) = 0$. Therefore, the root $\hat{\beta}_\phi$ of original estimating equation $U_\phi(\beta) = 0$ can be obtained by the iterative algorithm. Under regularity conditions, Jin *et al.* (2003) showed that each estimator obtained in the k iteration $\hat{\beta}_k$ is a legitimate estimator of β (consistent and asymptotically normal) when the initial estimator is consistent and asymptotically normal, and if the $\hat{\beta}_k$ converges as $k \to \infty$, the limit is indeed the solution of the $U_\phi(\beta) = 0$. Specifically, Jin *et al.* (2003) showed that $\hat{\beta}_k$ is asymptotically a weighted average of $\hat{\beta}_\phi$ and $\hat{\beta}_G$,

$$\hat{\beta}_k = \left\{\left(A_\phi + D_\phi\right)^{-1} D_\phi\right\}^k \hat{\beta}_G + \left[I - \left\{\left(A_\phi + D_\phi\right)^{-1} D_\phi\right\}^k\right]\hat{\beta}_\phi + o_p\left(n^{-\frac{1}{2}}\right), \tag{3.11}$$

where I is the $p \times p$ identity matrix,

$$D_\phi = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int \dot{\psi}_0(t) S^{(0)}(\beta_0; t) \left\{X_i - \frac{\sum_{j=1}^n X_j I\left\{e_j(\beta) \geq e_i(\beta)\right\}}{\sum_{j=1}^n I\left\{e_j(\beta) \geq e_i(\beta)\right\}}\right\}^{\otimes 2} dN_i(\beta_0; t), \tag{3.12}$$

and $\dot{\psi}_0(t)$ is the derivative of $\psi_0(t)$, the limit of $\psi(\beta_0; t)$ as $n \to \infty$.

3.3. Least-squares-type estimator

When there is no censoring, T_i, $i = 1, \ldots, n$ are completely observed, the classical least-squares estimator of β is obtained by minimizing $\sum_{i=1}^n (\log T_i - X_i^T\beta)^2$ in terms of β. The minimizer is the solution of the equation

$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(\log T_i - X_i^T \beta\right) = 0, \tag{3.13}$$

where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$.

In the presence of censoring, we can only observe Y_i and Equation (3.13) cannot be used directly. Buckly and James (1979) proposed to replace $\log T_i$ by following

$$\hat{Y}_i(\beta) = \delta_i \log Y_i + (1 - \delta_i)\left\{\frac{\int_{e_i(\beta)}^\infty u d\hat{F}_\beta(u)}{1 - \hat{F}_\beta\left(e_i(\beta)\right)} + X_i^T \beta\right\}, \tag{3.14}$$

where $\hat{F}_\beta(t)$ is the Kaplan-Meier estimator of $F(t)$ based on $\{e_i(\beta), \delta_i\}$ $(i = 1, \ldots, n)$, i.e.,

$$\hat{F}_\beta(t) = 1 - \prod_{i:e_i(\beta)<t} \left[1 - \frac{\delta_i}{\sum_{j=1}^n 1\left\{e_j(\beta) \geq e_i(\beta)\right\}}\right].$$

The resulting Buckly-James estimator is the solution to the follwoing equation:

$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(\hat{Y}_i(\beta) - X_i^T \beta\right) = 0. \tag{3.15}$$

The difficulty to solve the Equation (3.15) lies in that the function $\sum_{i=1}^{n}(X_i - \bar{X})(\hat{Y}_i(\beta) - X_i^T\beta)$ is neither continuous nor componentwise monotone in β. To overcome these difficulties, Jin *et al.* (2006) developed an iterative procedure to obtain a class of consistent and asymptotically normal estimators. Let $\bar{Y}(b) = n^{-1}\sum_{i=1}^{n}\hat{Y}_i(b)$, and define

$$U(\beta, b) = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(\hat{Y}_i(b) - X_i^T\beta\right), \tag{3.16}$$

or equivalently

$$U(\beta, b) = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left\{\hat{Y}_i(b) - \bar{Y}(b) - \left(X_i - \bar{X}\right)^T\beta\right\}. \tag{3.17}$$

Set $U(\beta, b) = 0$, which yields

$$\beta = Q(b) = \left\{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^{\otimes 2}\right\}^{-1}\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left\{\hat{Y}_i(b) - \bar{Y}(b)\right\}\right]. \tag{3.18}$$

It leads to an iterative algorithm

$$\hat{\beta}_{(m)} = Q\left(\hat{\beta}_{(m-1)}\right), \quad m \geq 1. \tag{3.19}$$

With the consistent initial estimator $\hat{\beta}_{(0)} = \hat{\beta}_G$, the Gehan-type estimator in the Section 3.1, Jin *et al.* (2006) showed that $\hat{\beta}_m$ is consistent and asymptotically normal for every $m \geq 1$, and $\hat{\beta}_{(m)}$ is asymptotically a weighted average of the Buckley-James estimator $\hat{\beta}_{BJ}$ and $\hat{\beta}_G$,

$$\hat{\beta}_m = \left(I - D^{-1}A\right)^m\hat{\beta}_G + \left[I - \left(I - D^{-1}A\right)^m\right]\hat{\beta}_{BJ} + o_p\left(n^{-\frac{1}{2}}\right), \tag{3.20}$$

where $D = \lim_{n\to\infty}(1/n)\sum_{i=1}^{n}(X_i - \bar{X})^{\otimes 2}$ and A is the slope matrix of the Buckley-James estimating function.

3.4. Kernel-smoothed profile likelihood estimator

Zeng and Lin (2007) proposed a kernel-smoothed profile likelihood estimator by an approximate nonparametric maximum likelihood method. Based on the semiparametric AFT model (2.1), the observed data (Y_i, δ_i, X_i), $i = 1, \ldots, n$ yields the following log-likelihood function of β and hazard function $\lambda(\cdot)$ of the error ϵ:

$$\text{loglik} = \frac{1}{n}\sum_{i=1}^{n}\left[-\delta_i X_i^T\beta + \delta_i\lambda\left(e_i(\beta)\right) - \Lambda\left(e_i(\beta)\right)\right], \tag{3.21}$$

Zeng and Lin (2007) showed that the direct maximization of the log-likelihood does not yield a solution. It also does not lead to the maximization of a profile likelihood obtained by replacing the $\lambda(e_i(\beta))$ by an estimator $\delta_i/\sum_{j=1}^{n} I\{e_j(\beta) \geq e_i(\beta)\}$ nor the maximization of a sieve profile likelihood obtained by replacing the $\lambda(\cdot)$ by a piecewise constant approximation. On the other hand, an approximation of

$$\frac{dP\left(\delta = 1, e(\beta) \leq t\right)/dt}{P\left(e(\beta) \geq t\right)}$$

by

$$\frac{1}{t}\frac{(na_n)^{-1}\sum_{i=1}^{n}\delta_i K\left(\frac{e_i(\beta)-\log t}{a_n}\right)}{\int_{\log(t)}^{\infty}(na_n)^{-1}\sum_{i=1}^{n}\delta_i K\left(\frac{e_i(\beta)-s}{a_n}\right)ds},$$

where $K(\cdot)$ is a kernel function (such as the standard normal density function) and a_n is a bandwidth, yields a kernel-smoothed approximation of the likelihood of β,

$$\begin{aligned} l_n^s(\beta) = &-\frac{1}{n}\sum_{i=1}^{n}\delta_i X_i^T\beta - \frac{1}{n}\sum_{i=1}^{n}\delta_i e_i(\beta) + \frac{1}{n}\sum_{i=1}^{n}\delta_i \log\left\{\frac{1}{na_n}\sum_{j=1}^{n}\delta_j K\left(\frac{e_j(\beta)-e_i(\beta)}{a_n}\right)\right\} \\ &-\frac{1}{n}\sum_{i=1}^{n}\delta_i\log\left\{\frac{1}{n}\sum_{j=1}^{n}\int_{-\infty}^{\frac{e_j(\beta)-e_i(\beta)}{a_n}}K(s)ds\right\}. \end{aligned} \tag{3.22}$$

The kernel-smoothed profile likelihood estimator $\hat{\beta}_{pl}$ is obtained by maximizing the $l_n^s(\beta)$ and its variance-covariance estimator is obtained by the use of the information matrix of the $l_n^s(\beta)$.

Given $\hat{\beta}_{pl}$, the $\lambda(\cdot)$ and $\Lambda(\cdot)$ can be estimated by

$$\hat{\lambda}_n(t) = \frac{(na_n t)^{-1}\sum_{i=1}^{n}\delta_i K\left(\frac{e_i(\hat{\beta}_{pl})-\log t}{a_n}\right)}{(n)^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\frac{e_i(\hat{\beta}_{pl})-\log t)}{a_n}}K(u)du}, \tag{3.23}$$

$$\hat{\Lambda}_n(t) = \int_{-\infty}^{\log t}\frac{(na_n)^{-1}\sum_{i=1}^{n}\delta_i K\left(\frac{e_i(\hat{\beta}_{pl})-\log t}{a_n}\right)}{(n)^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\frac{e_i(\hat{\beta}_{pl})-\log t}{a_n}}K(u)du}ds. \tag{3.24}$$

Under regularity conditions, Zeng and Lin (2007) showed that the estimator $\hat{\beta}_{pl}$ achieves semiparametric efficiency. The challenge for the practical use is the choice of the Kernel function, bandwidth, and the stability of the $l_n^s(\beta)$. Currently, there is no statistical software that implements the procedure.

4. Variance-covariance estimation

Except the kernel-smoothed profile likelihood estimator, the variance-covariance estimation of the Gehan-type rank estimator, weighted logrank-type estimator, and least-squares-type estimator is difficult because the corresponding variance-covariance matrices involve nonparametric estimation of the underlying probability density function. Here two computationally feasible approaches will be presented: resampling method and induced smoothing method.

4.1. Resampling method

The resampling method is similar to the resampling scheme similar to those in Rao and Zhao (1992), Parzen *et al.* (1994) and Jin *et al.* (2001). The basic idea is to perturb the objective functions that are minimized or estimating equations that are solved. Let Z_i $(i = 1, \ldots, n)$ be n independent positive random variables satisfying $E(Z_i) = \mathrm{Var}(Z_i) = 1$ and be independent of the observed data $(\tilde{T}_i, \delta_i, X_i)$ $(i = 1, \ldots, n)$. The Z_i can be generated by a number of known distributions, such as the standard exponential distribution or Poisson distribution Poisson(1).

4.1.1. Variance-covariance estimation for Gehan-type estimator

Jin *et al.* (2003) perturbed the objective function (3.2) and used following function

$$G_n^*(\beta) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i\left\{e_i(\beta) - e_j(\beta)\right\}^{-} Z_i. \tag{4.1}$$

The minimisation of G_n^* is also a linear programming problem and can be implemented with many existing softwares. Let $\hat{\beta}_G^*$ be the minimiser of G_n^*, it can be shown that $\hat{\beta}_G^*$ is the root of the estimating function

$$U_G^*(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i(X_i - X_j)I\Big\{e_i(\beta) - e_j(\beta) \leq 0\Big\}Z_i. \tag{4.2}$$

Jin *et al.* (2003) showed that the asymptotic distribution of $n^{1/2}(\hat{\beta}_G - \beta_0)$ can be approximated by the conditional distribution of $n^{1/2}(\hat{\beta}_G^* - \hat{\beta}_G)$ given the data $(\tilde{T}_i, \delta_i, X_i)$ $(i = 1, \ldots, n)$. Consequently, the distribution of $\hat{\beta}_G$, can be approximated by a large number of realizations of $\hat{\beta}_G^*$ by repeatedly generating the random sample $(Z_1, \ldots, Z_n)$ with the observed data $(\tilde{T}_i, \delta_i, X_i)$ $(i = 1, \ldots, n)$. The empirical variance-covariance matrix of $\hat{\beta}_G^*$ provides an estimate of the variance-covariance matrix of $\hat{\beta}_G$. The confidence intervals and hypothesis testing for individual β can be carried out based on the empirical distribution of $\hat{\beta}_G^*$ or by the Wald method.

The developed methods are implemented in the Splus/R package `lss` and the design and application features of `lss` are provided in Huang and Jin (2007).

4.1.2. Variance-covariance estimation for weighted logrank-type rank estimator

Due to the similarity between the objective function (3.10) for general weighted logrank-type rank estimator and the objective function (3.2) for Gehan-type estimator, the variance-covariance matrix of the $\hat{\beta}_{(k)}$ can be obtained by a similar resampling approach. Specifically, we can construct the perturbed objective function

$$L_\phi^*(\beta; b) := n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi(b; e_i(b))\delta_i\left\{e_i(\beta) - e_j(\beta)\right\}^{-} Z_i, \tag{4.3}$$

and define $\hat{\beta}_0^* = \hat{\beta}_G^*$ and $\hat{\beta}_k^* = \arg\min_\beta L_\phi^*(\beta; \hat{\beta}_{k-1}^*)$ $(k \geq 1)$. Jin *et al.* (2003) showed that the asymptotic distribution of $n^{1/2}(\hat{\beta}_k - \beta)$ can be approximted by the conditional distribution of $n^{1/2}(\hat{\beta}_k^* - \hat{\beta}_k)$ given the data $(\tilde{T}_i, \delta_i, X_i)$ $(i = 1, \ldots, n)$. Inference on β can then be carried out on the basis of the empirical distribution of $\hat{\beta}_k^*$.

4.1.3. Variance-covariance estimation for least-squares-type estimator

Jin *et al.* (2006) showed that the covariance matrix of $\hat{\beta}_m$ can be approximated by a resampling procedure. A theoretical derivation showed that both the Kaplan-Meier estimator used for the $\hat{Y}_i(b)$ and the estimating function (3.17) are required to be perturbed by the same Z_is, $(i = 1, \ldots, n)$. Specifically,

$$\hat{F}_b^*(t) = 1 - \prod_{i:e_i(b)<t}\left[1 - \frac{Z_i\delta_i}{\sum_{j=1}^{n} Z_j 1\Big\{e_j(b) \geq e_i(b)\Big\}}\right], \tag{4.4}$$

and

$$\hat{Y}_i^*(b) = \delta_i \tilde{Y}_i + (1-\delta_i)\left\{\frac{\int_{e_i(b)}^{\infty} u d\hat{F}_b^*(u)}{1-\hat{F}_b^*(e_i(b))} + X_i^T b\right\}, \tag{4.5}$$

$$Q^*(b) = \left\{\sum_{i=1}^n Z_i \left(X_i - \bar{X}\right)^{\otimes 2}\right\}^{-1}\left[\sum_{i=1}^n Z_i \left(X_i - \bar{X}\right)\left\{\hat{Y}_i^*(b) - \bar{Y}^*(b)\right\}\right]. \tag{4.6}$$

Then, $Q^*(b)$ leads to an iterative process $\hat{\beta}_m^* = Q^*(\hat{\beta}_{m-1}^*)$, $m \geq 1$. By setting the initial value $\hat{\beta}_0^* = \hat{\beta}_G^*$, the iteration procedure $\hat{\beta}_k^* = L^*(\hat{\beta}_{k-1}^*)$ yields a $\hat{\beta}_k^*$ $(1 \leq k \leq m)$. Jin *et al.* (2006) showed that the asymptotic distribution of $n^{1/2}(\hat{\beta}_k - \beta)$ can be approximted by the conditional distribution of $n^{1/2}(\hat{\beta}_k^* - \hat{\beta}_k)$ given the data $(\tilde{T}_i, \delta_i, X_i)$ $(i = 1, \dots, n)$. Therefore, by generating random samples of $(Z_1, \dots, Z_n)$ repeatedly N times, we can obtain N realizations of $\hat{\beta}_m^*$, denoted by $\hat{\beta}_{m,j}^*$ $(j = 1, \dots, N)$. For each $m \geq 1$, the covariance matrix of $\hat{\beta}_m$ can be estimated by

$$s^2 = \frac{1}{N-1}\sum_{j=1}^N \left(\hat{\beta}_{m,j}^* - \bar{\beta}_m^*\right)\left(\hat{\beta}_{m,j}^* - \bar{\beta}_m^*\right)^T, \tag{4.7}$$

where $\bar{\beta}_m^* = (1/N)\sum_{j=1}^N \hat{\beta}_{m,j}^*$.

The developed methods are implemented in the Splus/R package `lss` and the design and application features of `lss` are provided in Huang and Jin (2007).

4.2. Induced smoothing method

Brown and Wang (2005) proposed a general variance estimation procedure based on an induced smoothing for non-smooth estimating functions. The approach is computationally efficient and easy to implement when the smoothing in terms of integration with respect to a Gaussian distribution has an explicit form. For the Gehan-type rank estimator, the induced smoothing in Brown and Wang (2005) yields an explicit form and its varaince-covariance matrix can be easily estimated as shown in Brown and Wang (2007). However, the induced smoothing in Brown and Wang (2005) for the general logrank-weighted estimator and least-squares-type estimator does not yield an explicit form and cannot be used directly. To overcome the difficulty, Jin *et al.* (2015) developed a general Monte carlo approach for variance estimation for estimators based on induced smoothing which can be applied to both the general logrank-weighted estimator and least-squares-type estimator.

4.2.1. Induced smoothing for Gehan-type rank estimator

Brown and Wang (2007) and Johnson and Strawderman (2009) applied the induced smoothing method in Brown and Wang (2005) to the Gehan-type rank estimator $\hat{\beta}_G$. Because $\sqrt{n}(\hat{\beta}_G - \beta) \sim N(0, A_G^{-1} B_G A_G^{-1})$ as $n \to \infty$, Brown and Wang (2007) set $\beta = \hat{\beta}_G + n^{-1/2}[A_G^{-1} B_G A_G^{-1}]^{1/2} W$, where $W \sim N(0, I)$, and introduced following induced smoothing estimating function

$$\tilde{U}_n^G(V;\beta) = E_W U_n^G\left(\beta + n^{-\frac{1}{2}} V^{\frac{1}{2}} W\right), \tag{4.8}$$

where E_W denotes the expectation with respect to W. Brown and Wang (2007) showed that the expression (4.8) has an explicit closed form, specifically,

$$\tilde{U}_n^G(V;\beta) = \sum_{j=1}^{n} \delta_j \sum_{i=1}^{n} \left(X_i - X_j\right) \Phi\left(\frac{\sqrt{n}\left(e_i(\beta) - e_j(\beta)\right)}{\sqrt{\left(X_i - X_j\right)^T V \left(X_i - X_j\right)}} \right), \tag{4.9}$$

where $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$. The resulting induced smoothed estimating function $\tilde{U}_n^G(V;\beta)$ is continuous and differentiable function of β and $\tilde{U}_n^G(V;\beta) = 0$ can be solved iteratively with the initial values of $\hat{\beta}_G$ and $V_0 = I$. Therefore, they suggested to obtain $\hat{\beta}$ and $\hat{V}$ by jointly solving $\tilde{U}_n^G(V;\beta) = 0$ and the following equation

$$V = \tilde{A}^{-1}\left(V;\hat{\beta}\right) \hat{B}_G \tilde{A}^{-1}\left(V;\hat{\beta}\right), \tag{4.10}$$

where $\tilde{A}(V;\beta) = \partial \tilde{U}_n^G(V;\beta)/\partial\beta$ and $\hat{B}_G$ is a consistent estimator of B_G. Brown and Wang (2007) showed that the resulting parameter estimates and their covariance matrix usually converge in a few steps.

4.2.2. Induced smoothing for weighted logrank-type rank estimator

For the weighted logrank-type rank estimator $\hat{\beta}_\phi$, $n^{1/2}(\hat{\beta}_\phi - \beta)$ is asymptotically zero-mean normal with covariance matrix $A_\phi^{-1} B_\phi A_\phi^{-1}$. Ying *et al.* (1992) showed that

$$\hat{B}_\phi = \frac{1}{n}\sum_{i=1}^{n} \left\{\phi_i\left(\hat{\beta}\right)\right\}^2 \delta_i \left(\frac{\sum_{j=1}^{n} 1\left\{e_j\left(\hat{\beta}\right) \geq e_i\left(\hat{\beta}\right)\right\} X_j^{\otimes 2}}{\sum_{j=1}^{n} 1\left\{e_j\left(\hat{\beta}\right) \geq e_i\left(\hat{\beta}\right)\right\}} - \left[\frac{\sum_{j=1}^{n} 1\left\{e_j\left(\hat{\beta}\right) \geq e_i\left(\hat{\beta}\right)\right\} X_j}{\sum_{j=1}^{n} 1\left\{e_j\left(\hat{\beta}\right) \geq e_i\left(\hat{\beta}\right)\right\}} \right]^{\otimes 2} \right).$$

The challenging part is to estimate A_ϕ. To estimate A_ϕ, we can introduce following induced smoothing estimating function similar to (4.8),

$$\tilde{U}_\phi(V;\beta) = E_W U_\phi\left(\beta + n^{-\frac{1}{2}} V^{\frac{1}{2}} W\right), \tag{4.11}$$

where E_W denotes the expectation with respect to W. However, the expression (4.11) is too complicated to be written out in simple analytic forms. Consequently, the estimation of $\tilde{A}_\phi(V;\beta) = \partial\tilde{U}_\phi(V;\beta)/\partial\beta$ is difficult to obtain. Using Stein's identity, Jin *et al.* (2015) showed that

$$\tilde{A}_\phi(V;\beta) = n^{\frac{1}{2}} \left[E_W U_\phi\left(\beta + n^{-\frac{1}{2}} V^{\frac{1}{2}} W\right) W^T\right] V^{-\frac{1}{2}}. \tag{4.12}$$

Based on the result (4.12), Jin *et al.* (2015) showed a general Monte carlo approach that can be used to estimate $\tilde{A}_\phi(V;\beta)$ as following:

Step 1: Calculate a consistent estimator $\hat{B}_\phi$ of the matrix B_ϕ.

Step 2: Choose $V_0 = I$ and a large number m.

Step 3: For the k^{th} step ($k \geq 1$), generate z_j, $j = 1, \ldots, m$ from multivariate normal distribution $N(0, I)$. Estimate $\tilde{A}_\phi(V;\beta)$ by

$$A_k = A(V_{k-1}) = n^{\frac{1}{2}} \frac{1}{m} \sum_{j=1}^{m} U\left(\hat{\beta} + n^{-\frac{1}{2}} V_{k-1} z_j\right) z_j^T V_{k-1}^{-\frac{1}{2}} \tag{4.13}$$

Step 4: Calculate $V_k = A_k^{-1}\hat{D}A_k^{-1}$.

Step 5: Repeat Step 3 and Step 4 for next k until V_k converges.

The covariance matrix of $\sqrt{n}(\hat{\beta} - \beta)$ will be estimated using the V_k at the convergence in the above iterative algorithm.

In Jin *et al.* (2015), it is also shown that the Step 3 can be replaced by an approach based on Gaussian quadrature weights.

4.2.3. Induced smoothing for least-squares-type estimator

For the least-squares-type estimator $\hat{\beta}_m$, induced smoothing estimator can developed similar to the weighted logrank-type rank estimator in the Section 4.2.2. Here we present the results for the Buckley-James estimator $\hat{\beta}_{BJ}$ of the estimating equation (3.15). Ying *et al.* (1992) showed that $n^{1/2}(\hat{\beta}_{BJ} - \beta)$ is asymptotically zero-mean normal with covariance matrix $A_{BJ}^{-1}B_{BJ}A_{BJ}^{-1}$. Ying *et al.* (1992) also showed that

$$\hat{B}_{BJ} = \frac{1}{n}\int\left(\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^{\otimes 2} 1\left\{e_i\left(\hat{\beta}\right) \geq t\right\} - \frac{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)1\left\{e_i\left(\hat{\beta}\right) \geq t\right\}\right]^{\otimes 2}}{\sum_{i=1}^{n} 1\left\{e_i\left(\hat{\beta}\right) \geq t\right\}}\right)$$
$$\times \frac{\left(\int\left(1 - \hat{F}_{\hat{\beta}}(s)\right)1\{s \geq t\}ds\right)^2}{\left(1 - \hat{F}_{\hat{\beta}}(t)\right)^3} d\hat{F}_{\hat{\beta}}(t).$$

Again, the challenging part is to estimate A_{BJ}. To estimate A_{BJ}, we can introduce following induced smoothing estimating function similar to (4.8) and (4.11),

$$\tilde{U}_{BJ}(V;\beta) = E_W U_{BJ}\left(\beta + n^{-\frac{1}{2}}V^{\frac{1}{2}}W\right), \tag{4.14}$$

where E_W denotes the expectation with respect to W. The expression (4.14) again does not have a simple analytic form. Consequently, the estimation of $\tilde{A}_\phi(V;\beta) = \partial\tilde{U}_\phi(V;\beta)/\partial\beta$ can be done a Monte Carlo method similar to the one used for the expression (4.12), i.e.,

$$\tilde{A}_{BJ}(V;\beta) = n^{\frac{1}{2}}\left[E_W U_{BJ}\left(\beta + n^{-\frac{1}{2}}V^{\frac{1}{2}}W\right)W^T\right]V^{-\frac{1}{2}}, \tag{4.15}$$

see Jin *et al.* (2015). The remaining steps are straightforward as in the Section 4.2.2.

5. Other developments and remarks

In this paper, we have reviewed recently developed methods for point and variance estimation for the AFT model in the analysis of right censored data. The development is based on the assumption that errors are independent and identically distributed. The resampling approach offers valid inference but numerically intensive for large datasets, on the other hand, the induced smoothing approach is attractive due to its computational efficiency.

The paper of Jin and Ying (2004) studied asymptotic theory of rank estimation for AFT model under fixed censorship. Zhou (1992) and Jin (2007) studied M-estimation for the AFT models.

For the hypothesis testing framework, empirical likelihood approach has also been developed, see Zhou (2005a, 2015b), Zhou and Li (2008). For the heteroscedastic errors, Stute (1993, 1996) studied convergence properties of weighted estimators and Zhou *et al.* (2012) developed an empirical likelihood approach under a hypothesis testing framework.

For time-dependent covariates, there has been theoretical development, see Robins and Tsiatis (1992) and Lin and Ying (1995). The profile nonparametric likelihood approach of Zeng and Lin (2007) is applicable theoretically, but there has been no numerical investigation.

There are still many research problems, such as model checking, variable selection and efficient and reliable software development.

References

Buckley J and James I (1979). Linear regression with censored data, *Biometrika*, **66**, 429–436.

Brown BM and Wang YG (2005). Standard errors and covariance matrices for smoothed rank estimators, *Biometrika*, **92**, 732–746.

Brown BM and Wang YG (2007). Induced smoothing for rank regression with censored survival times, *Statistics in Medicine*, **26**, 828–836.

Cox DR (1972). Regression models and life tables (with discussion), *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

Cox DR and Oakes D (1984). *Analysis of Survival Data*, Chapman & Hall, London.

Fygenson M and Ritov Y (1994). Monotone estimating equations for censored data, *Annals of Statistics*, **22**, 732–746.

Gehan EA (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples, *Biometrika*, **52**, 203–223.

Huang L and Jin Z (2007). LSS: an S-Plus/R program for the accelerated failure time model to right censored data based on least-squares principle, *Computer Methods and Programs in Biomedicine*, **86**, 45–50.

Jin Z (2007). *M*-estimation in regression models for censored data, *Journal of Statistical Planning and Inference*, **137**, 3894–3903.

Jin Z, Lin DY, Wei LJ, and Ying Z (2003). Rank-based inference for the accelerated failure time model, *Biometrika*, **90**, 341–353.

Jin Z, Lin DY, and Ying Z (2006). On least-squares regression with censored data, *Biometrika*, **93**, 147–161.

Jin Z, Shao Y, and Ying Z (2015). A Monte Carlo method for variance estimation for estimators based on induced smoothing, *Biostatistics*, **16**, 179–188.

Jin Z and Ying Z (2004). Asymptotic theory in rank estimation for AFT model under fixed censorship, *Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life*, Editors: M. Nikulin, N. Balakrishnan, M. Mesbah, and N. Limnios, Boston: Birkhauser, 107–120.

Jin Z, Ying Z, and Wei LJ (2001). A simple resampling method by perturbing the minimand, *Biometrika*, **88**, 381–390.

Johnson LM and Strawderman RL (2009). Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data, *Biometrika*, **96**, 577–590.

Kalbfleisch JD and Prentice RL (2002). *The Statistical Analysis of Failure Time Data* (2nd ed), John Wiley, Hoboken.

Koul H, Susarla V, and Van Ryzin J (1981). Regression analysis with randomly right-censored data, *Annals of Statistics*, **9**, 1276–1288.

Lai TL and Ying Z (1991a). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data, *Annals of Statistics*, **19**, 1370–1402.

Lai TL and Ying Z (1991b). Rank regression methods for left-truncated and right-censored data,

Annals of Statistics, **19**, 531–556.
Lawless J (2003). *Statistical Models and Methods for Lifetime Data* (2nd ed), Wiley, New York.
Lin DY and Ying Z (1995). Semiparametric inference for the accelerated life model with time-dependent covariates, *Journal of Statistical Planning and Inference*, **44**, 47–63.
Mantel N (1966). Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Reports*, **50**, 163–170.
Miller R and Halpern J (1982).. Regression with censored data, *Biometrika*, **69**, 521–531.
Parzen MI, Wei LJ, and Ying Z (1994). A resampling method based on pivotal estimating functions, *Biometrika*, **81**, 341–50.
Powell JL (1984). Least absolute deviations estimation for the censored regression model, *Journal of Econometrics*, **25**, 303–325.
Prentice RL (1978). Linear rank tests with right censored data, *Biometrika*, **65**, 167–179.
Rao CR and Zhao LC (1992). Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap, *Sankhya A*, **54**, 323–331.
Ritov Y (1990). Estimation in a linear regression model with censored data, *Annals of Statistics*, **18**, 303–328.
Robins JM and Tsiatis AA (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates, *Biometrika*, **79**, 311–319.
Stute W (1993). Consistent estimation under random censorship when covariables are present, *Journal of Multivariate Analysis*, **45**, 89–103.
Stute W (1996). Distributional convergence under random censorship when covariables are present, *Scandinavian Journal of Statistics*, **23**, 461–471.
Tsiatis AA (1990). Estimating regression parameters using linear rank tests for censored data, *Annals of Statistics*, **18**, 354–372.
Wei LJ (1992). The accelerated failure time model: a usefu; alternative to the Cox regression model in survival analysis, *Statistics in Medicine*, **11**, 1871–1879.
Wei LJ, Ying Z, and Lin DY (1990). Linear regression analysis of censored survival data based on rank tests, *Biometrika*, **77**, 845–851.
Yang S (1997). Extended weighted log-rank estimating functions in censored regression, *Journal of the American Statistical Association*, **92**, 977–984.
Ying Z (1993). A large sample study of rank estimation for censored regression data, *Annals of Statistics*, **21**, 76–99.
Ying Z, Wei LJ, and Lin JS (1992). Prediction of survival probability based on a linear regression model, *Biometrika*, **79**, 205–209.
Zeng D and Lin DY (2007). Efficient estimation for the accelerated failure time model, *Journal of the American Statistical Association*, **102**, 1387–1396.
Zhou M (1992). M-estimation in censored linear models, *Biometrika*, **79**, 837–841.
Zhou M (2005a). Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model, *Biometrika*, **92**, 492–498.
Zhou M (2005b). Empirical likelihood ratio with arbitrarily censored/truncated data by a modified EM algorithm, *Journal of Computational and Graphical Statistics*, **14**, 643–656.
Zhou M, Kim M, and Bathke AC (2012). Empirical likelihood analysis for the heteroscedastic accelerated failure time model, *Statistica Sinica*, **22**, 295–316.
Zhou M and Li G (2008). Empirical likelihood analysis of the Buckley-James estimator, *Journal of Multivariate Analysis*, **99**, 649–664.

Received November 26, 2016; Revised November 27, 2016; Accepted November 27, 2016