Published in final edited form as:

J Korean Stat Soc. 2012 September 1; 41(3): 415–422. doi:10.1016/j.jkss.2012.01.003.

Semiparametric Estimation Methods for the Accelerated Failure Time Mixture Cure Model

Jiajia Zhang^a and Yingwei Peng^b

^aDepartment of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA

^bDepartment of Community Health and Epidemiology, Queen's University, Kingston, ON K7L 3N6, Canada

Abstract

This paper provides an overview of two semiparametric estimation methods recently proposed in the literature for the accelerated failure time mixture cure model. We prove that the two estimation methods are asymptotically equivalent. A simulation is conducted to investigate the rate of convergence of the two methods. We apply these methods to fit the accelerated failure time mixture cure model to the survival times of leukemia patients receiving bone marrow transplantation.

Keywords

Asymptotical equivalency; censored times; EM algorithm; M-estimation; rank-like estimation

1. Introduction

The accelerated failure time (AFT) model is one of the most important models in survival analysis. One advantage of the AFT model is that covariate effects on failure time are modeled directly, and interpretation of covariate effects may be simpler than in the proportional hazards model. Generally, estimates of the regression parameters in the model can be obtained through maximizing the likelihood function if the distribution of the error term in the model is specified (Lawless, 2003; Kalbfleisch and Prentice, 2002). Without a parametric assumption of the error term, Tsiatis (1990) proposed the rank estimation method and Ritov (1990) considered the general least square estimation method and proved the asymptotic equivalency between the rank estimation method and the general least square estimation method.

The common unstated assumption behind the AFT model is that all patients will eventually experience the event of interest, given that the follow-up time is long enough. However, with advances of medical treatments, patients can be cured of some diseases, and researchers are interested in assessing effects of a treatment or other covariates on the cure rate of the disease and on the failure time distribution of uncured patients. The mixture cure model, first introduced

by Boag (1949) and Berkson and Gage (1952), is an effective way to simultaneously model the cure rate of the treatment and the survival distribution of uncured patients.

Let T denote the failure time of interest, π be the probability of a patient being uncured, $S_u(t|\mathbf{x})$ be the survival probability of uncured patients, and $\mathbf{x}$ and $\mathbf{z}$ be observed values of two covariate vectors that may affect $S_u(t)$ and π respectively. The mixture cure model can be expressed as

$$S_{pop}(t|\mathbf{x}, \mathbf{z}) = \pi(\mathbf{z}) S_u(t|\mathbf{x}) + 1 - \pi(\mathbf{z}). \quad (1)$$

Usually, $\pi(\mathbf{z})$ is referred to as the “incidence” component and $S_u(t|\mathbf{x})$ is referred to as the “latency” component of the model. Effects of covariates on the cure rate in the incidence component are often assumed to be in the logistic link form:

$$\pi(\mathbf{z}) = \frac{\exp(\mathbf{bz})}{1 + \exp(\mathbf{bz})}, \quad (2)$$

where $\mathbf{b}$ is a row vector of unknown parameters. As mentioned in Price and Manatunga (2001), other link functions can also be considered. Examples include the probit link $\pi(\mathbf{z}) = \Phi(\mathbf{bz})$ and the complementary log-log link $\pi(\mathbf{z}) = 1 - \exp(-\exp(\mathbf{bz}))$. For the latency component, we consider the AFT model to describe effects of covariates on the failure time of uncured patients:

$$\log T = \beta\mathbf{x} + \varepsilon \quad \text{or} \quad S_u(t|\mathbf{x}) = S(\log(t) - \beta\mathbf{x}), \quad (3)$$

where β is a row vector of unknown parameters, and ε denotes the error term with an unknown survival function $S(\cdot)$. We refer to this model as the AFT mixture cure (AFTMC) model.

To estimate parameters in the AFTMC model, Li and Taylor (2002) first developed a semiparametric estimation method based on the EM algorithm and the M-estimator. Zhang and Peng (2007) recently proposed an alternative semiparametric approach by incorporating a rank-like estimation equation in the EM algorithm. However, the relationship between the two estimation methods is not clear. In this paper, we will reveal an asymptotic equivalency relationship between the two methods, and further compare the rate of convergence of the methods. The paper is organized as follows. We will first review the two semiparametric estimation methods in section 2. We then show in section 3 that the two methods are asymptotically equivalent. From the prospect of practice, we investigate two estimation methods with finite samples by considering the accuracy of estimates and the computational time for each method in section 4. We demonstrate an application of the model and estimation methods in section 5. Conclusions are given in section 6.

2. Semiparametric Estimation Methods

Let $\mathbf{O}$ be the set of observed data $\{(t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i), i = 1, \dots, n\}$, where t_i and δ_i are the observed survival time and censoring indicator (1 if t_i is uncensored and 0 otherwise), and $\mathbf{x}_i$ and $\mathbf{z}_i$ are the observed values of covariates $\mathbf{x}$ and $\mathbf{z}$ for the subject i . To fit data with the AFTMC model specified in (1), (2) and (3), we consider the EM algorithm to obtain the maximum likelihood estimates of unknown parameters in the model, denoted as $\Theta = \{\mathbf{b}, \beta, S(t)\}$. Define a latent random variable y_i with $y_i = 1$ if the i th individual is uncured and $y_i = 0$ if cured. Given values of $\mathbf{y} = (y_1, \dots, y_n)$, the complete log likelihood is $l_c(\mathbf{b}, \beta, S(\cdot); \mathbf{O}, \mathbf{y}) = l_{c1}(\mathbf{b}; \mathbf{O}, \mathbf{y}) + l_{c2}(\beta, S(\cdot); \mathbf{O}, \mathbf{y})$, where

$$l_{c1}(\mathbf{b}; \mathbf{O}, \mathbf{y}) = \sum_{i=1}^n y_i \log[\pi(\mathbf{z}_i)] + (1 - y_i) \log[1 - \pi(\mathbf{z}_i)],$$

$$l_{c2}(\beta, S(\cdot); \mathbf{O}, \mathbf{y}) = \sum_{i=1}^n y_i \delta_i \log[h(\log(t_i) - \beta \mathbf{x}_i)] + y_i \log[S(\log(t_i) - \beta \mathbf{x}_i)],$$

and $h(\cdot)$ is the corresponding hazard function of $S(\cdot)$.

The E-step computes the conditional expectation of the complete log-likelihood with respect to the latent variable y_i given the current estimates $\Theta^{(m)} = \{\mathbf{b}^{(m)}, \boldsymbol{\beta}^{(m)}, S^{(m)}(t)\}$ and observed data $\mathbf{O}$. Let $\omega_i^{(m)}$ denote $E(y_i | \Theta^{(m)}, \mathbf{O})$. It is easy to show that

$$\omega_i^{(m)} = \delta_i + (1 - \delta_i) \frac{\pi(\mathbf{z}_i) S(\log(t_i) - \beta \mathbf{x}_i)}{1 - \pi(\mathbf{z}_i) + \pi(\mathbf{z}_i) S(\log(t_i) - \beta \mathbf{x}_i)} \Big|_{\Theta = \Theta^{(m)}}, \quad (4)$$

and $E[l_c(\mathbf{b}, \boldsymbol{\beta}, S(\cdot); \mathbf{O}, \mathbf{y}) | \Theta^{(m)}, \mathbf{O}] = l_{c1}(\mathbf{b}) + l_{c2}(\boldsymbol{\beta}, S(\cdot))$, where

$$l_{c1}(\mathbf{b}) = \sum_{i=1}^n \omega_i^{(m)} \log[\pi(\mathbf{z}_i)] + (1 - \omega_i^{(m)}) \log[1 - \pi(\mathbf{z}_i)], \quad (5)$$

$$l_{c2}(\beta, S(\cdot)) = \sum_{i=1}^n \omega_i^{(m)} \delta_i \log[h(\log(t_i) - \beta \mathbf{x}_i)] + \omega_i^{(m)} \log[S(\log(t_i) - \beta \mathbf{x}_i)]. \quad (6)$$

The M-step is to maximize (5) and (6) with respect to unknown parameters $\mathbf{b}$, $\boldsymbol{\beta}$ and $S(\cdot)$. Maximizing $l_{c1}(\mathbf{b})$ in (5) with respect to $\mathbf{b}$ can be easily carried out using the Newton-Raphson algorithm, and maximizing $l_{c2}(\boldsymbol{\beta}, S(\cdot))$ in (6) with respect to $\boldsymbol{\beta}$ and $S(\cdot)$ can be carried out by the two methods described below.

2.1. Li and Taylor's Method

Li and Taylor (2002) proposed to maximize $l_{c2}(\boldsymbol{\beta}, S(\cdot))$ in (6) with respect to $\boldsymbol{\beta}$ by solving the following general estimating equation

$$\Phi(\boldsymbol{\beta}; g(\cdot)) = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) \left[\omega_i^{(m)} \delta_i g(\varepsilon_i) - \omega_i^{(m)} (1 - \delta_i) \frac{\int_{\varepsilon_i}^{\infty} g(v) dS(v)}{S(\varepsilon_i)} \right] = 0, \quad (7)$$

where $\bar{\mathbf{x}} = (n^{-1} \sum_{i=1}^n \mathbf{x}_i)$ and $g(\cdot)$ is an unspecified function. To cope with the unknown survival distribution $S(\cdot)$, Li and Taylor (2002) proposed to replace it by an estimator $S^{\boldsymbol{\beta}}(\cdot)$ based on the residuals $\log(t_i) - \boldsymbol{\beta}^{(m)} \mathbf{x}_i$, where $\boldsymbol{\beta}^{(m)}$ is the most current estimate of $\boldsymbol{\beta}$ (Taylor, 1995). Let $\tau_1 < \tau_2 < \dots < \tau_k$ be the distinct uncensored failure residuals, d_{τ_j} denote number of the uncensored failure residuals at τ_j , and $R(\tau_j)$ denote the risk set at τ_j . An estimate of $S(\cdot)$ in the current M-step is

$$S^{\beta}(\varepsilon)=\exp\left(-\sum_{j:\tau_j<\varepsilon}\frac{d\tau_j}{\sum_{i\in R(\tau_j)}\omega_i^{(m)}}\right) \quad (8)$$

(we require that $S^{\beta}(\varepsilon) = 0$ if $\varepsilon > \tau_k$). The g -function in (7) is replaced by the identity function and the bounded function, where the identity function $g(u) = u$ corresponds to the standard normal distribution of ε . When one such simplified g -functions is used and $S(\cdot)$ is replaced with its estimator $S^{\beta}(\cdot)$ in $\Phi(\beta, g(\cdot))$, the estimate of β can be obtained through solving (7).

2.2. Rank-like Estimation Method

Zhang and Peng (2007) recently proposed to maximize $l_{c2}(\beta, S(\cdot))$ in (6) by treating it as a log-likelihood function from a general AFT model

$$\log(T_i)=\beta\mathbf{x}_i+\varepsilon_i^*,$$

where the hazard function of ε_i^* is $\omega_i^{(m)}h(\varepsilon_i^*)$. Thus, the rank-like estimation method for the AFT model can be employed here to estimate β , and the corresponding estimating equation is

$$\Psi(\beta;k(\cdot))=\sum_{i=1}^n\delta_i k(\varepsilon_i^*)\left(\mathbf{x}_i-\frac{\sum_{j=1}^n\mathbf{x}_j\omega_j^{(m)}I(\varepsilon_j^*\geq\varepsilon_i^*)}{\sum_{j=1}^n\omega_j^{(m)}I(\varepsilon_j^*\geq\varepsilon_i^*)}\right), \quad (9)$$

where $k(\cdot)$ is a (predictable) weight function. An example of the weight function is the Gehan-

type weight function, $k(u)=\sum_{j=1}^n I(\varepsilon_j^*\geq u)\omega_j^{(m)}/n$. If this weight function is used, the estimating equation (9) can be simplified as

$$\Psi(\beta)=n^{-1}\sum_{i=1}^n\sum_{j=1}^n\delta_i(\mathbf{x}_i-\mathbf{x}_j)\omega_j^{(m)}I(\varepsilon_j^*\geq\varepsilon_i^*). \quad (10)$$

It is easy to show that this estimating function is monotone. Thus, if a solution of $\Psi(\beta) = 0$ exists, it will be unique and consistent. The asymptotic normality of $\Psi(\beta; k(\cdot))$ comes from the fact that it can be rewritten as a U-statistic with a symmetric kernel:

$$\Psi(\beta)=n^{-1}\sum_{i=2}^n\sum_{j=1}^{i-1}(\mathbf{x}_i-\mathbf{x}_j)[\delta_i\omega_i^{(m)}I(\varepsilon_i^*<\varepsilon_j^*)-\delta_j\omega_j^{(m)}I(\varepsilon_j^*<\varepsilon_i^*)].$$

The estimating function (10) can be taken as the gradient of a convex function

$$L_G(\beta)=n^{-1}\sum_{i=1}^n\sum_{j=1}^n\delta_i\omega_j^{(m)}|\varepsilon_i^*-\varepsilon_j^*|I(\varepsilon_i^*<\varepsilon_j^*). \quad (11)$$

Hence finding the root of $\Psi(\beta) = 0$ is equivalent to minimizing this convex function, which can be conveniently carried out by the linear programming method.

Given the nonparametric estimate of β from above procedures, an estimator, $S^{\beta}(\cdot)$, can be obtained based on the residuals $\log(t_i) - \beta^{(m)} \mathbf{x}_i$ as in (8).

Once $\beta^{(m)}$ and $S^{\beta}(\cdot)$ are obtained, the EM algorithm can proceed to the next E-step by updating $\omega_i^{(m)}$ in (4).

3. Asymptotical Equivalency

Ritov (1990) established the asymptotic equivalency between an M-estimator and a linear rank estimator for the semiparametric AFT model based on the martingale representation of the two estimators. The argument can be extended to the semiparametric AFTMC model by showing the asymptotic equivalency between the estimation method proposed in Li and Taylor (2002) and rank-like estimation method in Zhang and Peng (2007). This is achievable because the estimating equations in the M-step of the EM algorithm for the AFTMC model are similar to those in the AFT model. The major difference in the rank-like estimation method and Li and Taylor's method is in the estimating equations (7) and (9) in the M-step in the EM algorithm. The estimating equation for $\mathbf{b}$ based on (5) remains unchanged. Therefore, to show the asymptotic equivalency of the two estimation methods, we only need to show the asymptotic equivalency of the two estimating equations (7) and (9).

To present the main results of this paper, we change our notation slightly as in Ritov (1990). Let T_i and C_i be the true failure time and the censoring time respectively,

$\varepsilon_i^{\beta} = T_i - \beta' \mathbf{x}_i$ and $\zeta_i^{\beta} = C_i - \beta' \mathbf{x}_i$, and $F_n^{\beta}(\cdot)$ be an estimate of the baseline distribution function $1 - S(\cdot)$ given an estimate of β , which can be obtained from the estimate of $S(t)$ given (8) based on ε_i^{β} and ζ_i^{β} . A profile estimating equation of β can be written as

$$\Phi_n(\beta; g(\cdot)) = n^{-1/2} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) \left[\omega_i^{(m)} \delta_i g(\varepsilon_i^{\beta}) + \omega_i^{(m)} (1 - \delta_i) \frac{\int_{\zeta_i^{\beta}}^{\infty} g(u) dF_n^{\beta}(u)}{I - F_n^{\beta}(\zeta_i^{\beta})} \right], \quad (12)$$

Similarly, (9) can be written as

$$\Psi_n(\beta; k(\cdot)) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{\zeta_i^{\beta}} k(u) \left(\mathbf{x}_i - \frac{\sum_{j=1}^n \mathbf{x}_j \omega_j^{(m)} I(\varepsilon_j^{\beta} \wedge \zeta_j^{\beta} \geq u)}{\sum_{j=1}^n \omega_j^{(m)} I(\varepsilon_j^{\beta} \wedge \zeta_j^{\beta} \geq u)} \right) dN_i^{\beta}(u), \quad (13)$$

where $N_i^{\beta}(u) = I(\varepsilon_i^{\beta} \leq u, \delta_i = 1)$. We now add a term that is identical to 0 to the equation above to obtain

$$\Psi_n(\beta; k(\cdot)) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{\zeta_i^{\beta}} k(u) \left(\mathbf{x}_i - \frac{\sum_{j=1}^n \mathbf{x}_j \omega_j^{(m)} I(\varepsilon_j^{\beta} \wedge \zeta_j^{\beta} \geq u)}{\sum_{j=1}^n \omega_j^{(m)} I(\varepsilon_j^{\beta} \wedge \zeta_j^{\beta} \geq u)} \right) \{dN_i^{\beta}(u) - \omega_i^{(m)} I(\varepsilon_i^{\beta} \geq u) d\Lambda(u)\},$$

for any hazard function $\Lambda(\cdot)$.

Then we will represent $\Phi_n(\beta; g(\cdot))$ and $\Psi_n(\beta; k(\cdot))$ by the martingale notation. Let

$$Q_g(t; \varepsilon | F) = g(\varepsilon) I(\varepsilon \leq t) + \frac{\int_t^\infty g(u) dF(u)}{1 - F(t)} I(\varepsilon > t).$$

It is easy to show that $Q_g(t; \varepsilon | F)$ is a martingale. We have

$$\Phi_n(\beta; g(\cdot)) = n^{-1/2} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) \omega_i^{(m)} Q_g(\zeta_i^\beta; \varepsilon_i^\beta | F_n^\beta). \quad (14)$$

Let ω be the probability of being uncured. Because $\omega I(\varepsilon \leq t) = I(\varepsilon \leq t)$, (Note: If $\varepsilon \leq t$ is true, the subject is uncured and $\omega = 1$ following the definition. Thus $\omega I(\varepsilon \leq t) = 1$. If $\varepsilon \leq t$ is false, $\omega I(\varepsilon \leq t) = 0$. Therefore $\omega I(\varepsilon \leq t) = I(\varepsilon \leq t)$), we can define a counting process, which is a martingale

$$M(t; \varepsilon, \omega | F) = \left[I(\varepsilon \leq t) - \int_{-\infty}^t \omega I(\varepsilon \geq u) \frac{dF(u)}{1 - F(u-)} \right].$$

Then we have

$$\Psi_n(\beta; k(\cdot)) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{\zeta_i^\beta} k(u) \left(\mathbf{x}_i - \frac{\sum_{j=1}^n \mathbf{x}_j \omega_j^{(m)} I(\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u)}{\sum_{j=1}^n \omega_j^{(m)} I(\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u)} \right) dM(u; \varepsilon_i^\beta, \omega_i^{(m)} | F), \quad (15)$$

for any distribution function F .

We next show a relationship between $Q_g(\cdot, \varepsilon | F)$ and $M(\cdot, \varepsilon, \omega | F)$. For any distribution F on the real line, define the transformation W_F and L_F as follows

$$\begin{aligned} W_F g(t) &= g(t) - \frac{\int_t^\infty g(u) dF(u)}{1 - F(t)}, \\ L_F k(t) &= k(t) - \int_{-\infty}^t k(u) \frac{dF(u)}{1 - F(u-)}. \end{aligned}$$

Let $L_2(F)$ be the space of functions which are square integrable with respect to the distribution F . We have the following proposition similar to Proposition 3.1 in Ritov (1990):

Proposition 1: Suppose $g(t) \in L_2(F)$ and $\int g(t) dF(t) = 0$, then the martingale $Q_g(\cdot; \varepsilon | F)$ is related to the counting process martingale $M(\cdot; \varepsilon | F)$ by

$$\omega Q_g(t; \varepsilon | F) = \int_{-\infty}^t W_F g(u) dM(u; \varepsilon, \omega | F).$$

Proof: Following the proof of Proposition 3.1 in Ritov (1990), we have

$$\int_{-\infty}^t W_F g(u) \frac{dF(u)}{1-F(u-)} = \frac{\int_t^{\infty} g(u) dF(u)}{1-F(t)}.$$

Hence

$$\int_{-\infty}^t W_F g(u) dM(u; \varepsilon, \omega | F) = g(\varepsilon) I(\varepsilon \leq t) - \frac{\int_{\varepsilon}^{\infty} g(u) dF(u)}{1-F(\varepsilon)} I(\varepsilon \leq t) - \int_{-\infty}^{t \wedge \varepsilon} W_F g(u) \omega \frac{dF(u)}{1-F(u-)}.$$

Because $\omega I(\varepsilon \leq t) \equiv I(\varepsilon \leq t)$, we can rewrite above equation as

$$\begin{aligned} & \int_{-\infty}^t W_F g(u) dM(u; \varepsilon | F) \\ &= g(\varepsilon) I(\varepsilon \leq t) \omega \\ & \quad - \frac{\int_{\varepsilon}^{\infty} g(u) dF(u)}{1-F(\varepsilon)} I(\varepsilon \leq t) \omega - \omega \int_{-\infty}^{t \wedge \varepsilon} W_F g(u) \frac{dF(u)}{1-F(u-)} \\ &= \omega g(\varepsilon) I(\varepsilon \leq t) \\ & \quad + \omega \frac{\int_t^{\infty} g(u) dF(u)}{1-F(t)} I(\varepsilon > t) \\ &= \omega Q_g(t; \varepsilon | F) \end{aligned}$$

$$\text{Thus } \omega Q_s(t, \varepsilon | F) = \int_{-\infty}^t W_F g(u) dM(u; \varepsilon, \omega | F).$$

The following proposition establish the correspondence between $\Phi_n(\cdot)$ and $\Psi_n(\cdot)$.

Proposition 2: For all $\beta \in R_m$ and $g(\cdot), k(\cdot): R \rightarrow R$ and w.p.1,

$$\Phi_n(\beta; g(\cdot)) = \Psi_n(\beta; W_{F_n^\beta} g(u)) \quad (16)$$

$$\Psi_n(\beta; k(\cdot)) = \Phi_n(\beta; L_{F_n^\beta} k(\cdot)) \quad (17)$$

Proof: Following Proposition 1, we have

$$\Phi_n(\beta; g(\cdot)) = n^{-1/2} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) \left[\int_{-\infty}^{s_i^\beta} W_{F_n^\beta} g(u) dM(u; \varepsilon_i^\beta, \omega_i^{(m)} | F_n^\beta) \right].$$

Suppose that F_n^β has atoms only on uncensored observations and

$$\frac{dF_n^\beta(u)}{1-F_n^\beta(u-)} = \frac{\sum_{i=1}^n \delta_i I(u = \varepsilon_i^\beta)}{\sum_{j=1}^n I(\varepsilon_i^\beta \wedge s_i^\beta \geq u) \omega_j^{(m)}}.$$

Hence, for any function $h(\cdot)$,

$$\sum_{i=1}^n \int_{-\infty}^{\zeta_i^\beta} h(u) dM(u; \varepsilon_i^\beta, \omega_i^{(m)} | F_n^\beta) = \sum \delta_i h(\varepsilon_i^\beta) - \int \sum_{i=1}^n I(\varepsilon_i^\beta \wedge \zeta_i^\beta \geq u) \omega_i^{(m)} h(u) \frac{dF_n^\beta(u)}{1 - F_n^\beta(u-)} = 0 \text{ w.p. } 1.$$

Let

$$h(u) = \frac{\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) \omega_i^{(m)} I(\varepsilon_i^\beta \wedge \zeta_i^\beta \geq u)}{\sum_{i=1}^n \omega_i^{(m)} I(\varepsilon_i^\beta \wedge \zeta_i^\beta \geq u)} W_{F_n^\beta} g(u).$$

According to (18), we have

$$\sum_{i=1}^n \int_{-\infty}^{\zeta_i^\beta} \frac{\sum_{j=1}^n \mathbf{x}_j \omega_j^{(m)} I(\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u)}{\sum_{j=1}^n \omega_j^{(m)} I(\varepsilon_j^\beta \wedge \zeta_j^\beta \geq u)} W_{F_n^\beta} g(u) dM(u; \varepsilon_i^\beta, \omega_i^{(m)} | F_n^\beta) = \sum_{i=1}^n \int_{-\infty}^{\zeta_i^\beta} \bar{\mathbf{x}} W_{F_n^\beta} g(u) dM(u; \varepsilon_i^\beta, \omega_i^{(m)} | F_n^\beta)$$

Then we can obtain

$$\Psi_n(\beta; W_{F_n^\beta} g(u)) = n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{\zeta_i^\beta} (\mathbf{x}_i - \bar{\mathbf{x}}) W_{F_n^\beta} g(u) dM(u; \varepsilon_i^\beta | F_n^\beta) = \Phi_n(\beta; g(\cdot)).$$

Similar to Ritov (1990), we can show $\Psi_n(\beta; k(\cdot)) = \Phi_n(\beta; L_{F_n^\beta} k(\cdot))$.

The asymptotic equivalency between $\Psi(\beta; k(\cdot))$ and $\Phi(\beta; g(\cdot))$ follows immediately based on the correspondence, and thus the two estimation methods are asymptotically equivalent.

It is worthwhile pointing out that the asymptotic equivalency of these two methods is based on choices of M-estimators in Li and Taylor's method and the weight function in the rank-like estimation method. It is also obvious that the equivalency of the two methods does not depend on choices of the link function, $\pi(\cdot)$, as long as the same link function is used.

4. Simulation

We conduct a simulation study to understand the asymptotic equivalency of the two methods. In particular, we investigate the sample sizes required to achieve the equivalency. We generate 1000 simulated data sets with sample size of 200, 500 and 1000 from an AFTMC model. Each data set contains two covariates, one is a binary variable taking 0 or 1 with equal probability, and the other is a continuous variable drawn from a standard normal distribution. The covariate vector $\mathbf{z}$ in the AFTMC model contains the binary covariate only while the covariate $\mathbf{x}$ contains both the binary and continuous covariates. We assume $\beta = (\beta_1, \beta_2) = (1, 1)$ and $\mathbf{b} = (b_0, b_1) = (2, -1)$. The error term is generated from a standard normal distribution. The censoring is generated from the uniform distribution with an approximate 20% censoring rate. We fit the data sets with the AFTMC model using Li and Taylor's method with the identity function and the rank-like estimation method with the Gehan-type weight described in previous sections, and calculate the biases, and variances of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{b}_0$ and $\hat{b}_1$. We also compared the two estimation methods in terms of computing time required to complete the entire simulation computation of the 1000 data sets in each setting. The results are shown in Table 1.

The results demonstrate that when the sample size is as large as 1000, the two estimation methods show considerable agreement in their results, as the asymptotic equivalency suggests. When the sample size is smaller, there are noticeable differences in the results; particularly, the rank-like estimation method tends to have smaller biases in $\hat{\beta}_1$ and $\hat{\beta}_2$ than Li and Taylor's method. Moreover, the rank-like estimation method is about 4 times faster than Li and Taylor's method.

5. Real Data Analysis

As an illustration of the AFTMC model and the estimation methods discussed in previous sections, we consider a study of the bone marrow transplantation for leukemia patients discussed in Copelan et al. (1991). There are 137 patients in this study. The survival times from transplantation to relapse due to leukemia are recorded for all patients, some of which are censored. An important risk factor for the survival time is the type of the leukemia at the time of transplantation, which includes acute lymphoblastic leukemia (ALL) (38 patients), acute myelocytic leukemia (AML) low risk (54 patients), and AML high risk (45 patients). It is known that there are long term survivors among leukemia patients receiving bone marrow transplantation, and they can be treated as cured. This is also demonstrated by the plateau at the tails of the Kaplan-Meier survival curves of the three risk groups in Figure 1.

We also drew the logarithm of cumulative hazard functions of uncured patients in the three groups in Figure 2 to demonstrate that these curves are far from parallel, which suggests that cure models based on the proportional hazards assumption, such as the one in Peng and Dear (2000) and Sy and Taylor (2000), are inappropriate. Therefore we are interested in modeling the survival times by the AFTMC model, which can exam the effects of the leukemia type on the cure rate as well as the failure time of uncured patients. We fit the model using the two methods discussed in previous sections (with the same weight as in the simulation study) and the results are presented in Table 2. The standard errors of estimates are obtained using the bootstrap method based on 500 bootstrap samples. The table also shows the estimated cure rates for the three risk groups based on $\hat{\mathbf{b}}$.

The sample size of the data is 137. Per the results in the simulation, it is not surprising to see some differences in the estimated effects in the AFTMC model from the two estimation methods. The most important difference is that the AML-low risk group has a significantly higher cure rate and a longer survival time for uncured patients in the rank-like estimation method, while in Li and Taylor's method its effect on the survival time of uncured patients is not significant. The resulting cure rates in the three groups are different in the two methods even though both methods show that the AML-low risk group has the largest cure rate and the AML-low risk group has the smallest cure rate. The estimated survival function $S_{pop}(t)$ in (1) are presented in Figure 1 to better reveal the differences between the two estimation methods.

6. Conclusions

In this paper, we establish the asymptotic equivalence between the Li and Taylors's method and rank-like estimation method. In spite of the equivalence, we recommend the rank-like estimation method with Gehan type weight in practice due to its accuracy and the shorter computational time.

Acknowledgments

The authors thank the reviewers and associate editor for insightful comments, which have led to a great improvement in the presentation of the work. The project is supported by an award from the National Cancer Institute (NCI, Award Number R03CA137790).

References

- Berkson J, Gage RP. Survival curve for cancer patients following treatment. *J. Amer. Statist. Assoc.* 1952; 47:501–515.
- Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. Roy. Statist. Soc. Ser. B.* 1949; 11:15–44.
- Copelan EA, Biggs JC, Thompson JM, Crilley P, Szer J, Klein JP, Kapoor N, Avalos BR, Cunningham I, Atkinson K, Downs K, Harmon GS, Daly MB, Brodsky I, Bulova SI, Tutschka PJ. Treatment for Acute Myelocytic Leukemia with Allogeneic Bone Marrow Transplantation Following Preparation with Bu/Cy. *Blood.* 1991; 78:838–843. [PubMed: 1859895]
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data.* Hoboken, NJ: John Wiley & Sons; 2002.
- Klein, JP.; Moeschberger, ML. *Survival Analysis: Techniques for Censored and Truncated Data.* New York: Springer-Verlag Inc; 1997.
- Lawless, JF. *Statistical Models and Methods for Lifetime Data*, 2nd Edition. Hoboken, NJ: John Wiley & Sons; 2003.
- Li C-S, Taylor JMG. A semi-parametric accelerated failure time cure model. *Statist. Med.* 2002; 21(21): 3235–3247.
- Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics.* 2000; 56:237–243. [PubMed: 10783801]
- Price DL, Manatunga AK. Modelling survival data with a cured fraction using frailty models. *Statist. Med.* 2001; 20(9–10):1515–1527.
- Ritov Y. Estimation in a linear regression model with censored data. *Ann. Statist.* 1990; 18(1):303–328.
- Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics.* 2000; 56:227–236. [PubMed: 10783800]
- Taylor JMG. Semi-parametric estimation in failure time mixture models. *Biometrics.* 1995; 51:899–907. [PubMed: 7548707]
- Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* 1990; 18(1):354–372.
- Zhang J, Peng Y. A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statist. Med.* 2007; 26(16):3157–3171.

Figure 1. Estimated survival curves from three groups based on Kaplan-Meier, Li and Taylor's, and rank-like estimation methods respectively.

Figure 2. Estimated log cumulative hazard functions of the three risk groups based on the Kaplan-Meier method.

Table 1

Biases and empirical variance of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{b}_0$, $\hat{b}_1$ of 1000 simulated data sets from the rank-like estimation method and Li and Taylor's method

<i>n</i>	Parameter	Rank-like		Li & Taylor	
		Bias	Variance	Bias	Variance
200	$\hat{\beta}_1$	0.002	0.029	0.064	0.016
	$\hat{\beta}_2$	-0.001	0.071	-0.024	0.018
	$\hat{b}_0$	0.023	0.111	0.024	0.111
	$\hat{b}_1$	0.007	0.165	-0.006	0.166
Computing time (seconds)		2803		10219	
500	$\hat{\beta}_1$	0.003	0.010	0.041	0.006
	$\hat{\beta}_2$	-0.003	0.003	-0.012	0.007
	$\hat{b}_0$	0.017	0.040	0.015	0.040
	$\hat{b}_1$	-0.011	0.061	-0.012	0.061
Computing time (seconds)		5795		23355	
1000	$\hat{\beta}_1$	0.0004	0.005	0.0002	0.003
	$\hat{\beta}_2$	0.003	0.001	-0.003	0.001
	$\hat{b}_0$	0.008	0.020	0.008	0.020
	$\hat{b}_1$	-0.003	0.031	-0.002	0.031
Computing time (seconds)		16692		69321	

Table 2

Estimated parameters and standard errors from the AFTMC model by the EM algorithm with rank estimation method and Li and Taylor's method.

Parameter	Rank-like		Li and Taylor	
	Estimate	Standard error	Estimate	Standard error
<i>Latency part:</i>				
ALL (vs AML-high risk)	0.351	0.396	0.321	0.329
AML-low risk (vs AML-high risk)	0.910	0.276	0.084	0.389
<i>Incidence part:</i>				
Intercept	1.135	0.393	1.579	0.496
ALL (vs AML-high risk)	-0.497	0.511	-0.233	0.831
AML-low risk (vs AML-high risk)	-1.178	0.534	-1.548	0.609
<i>Cure rate:</i>				
ALL		0.346		0.207
AML-low risk		0.511		0.492
AML-high risk		0.243		0.171