Statistica Sinica Preprint No: SS-2016-0231R2

Title	Smoothed Rank Regression for the Accelerated Failure Time Competing Risks Model with Missing Cause of Failure
Manuscript ID	SS-2016-0231.R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0231
Complete List of Authors	Alan Wan Zhiping Qiu Yong Zhou and Peter Gilbert
Corresponding Author	Alan Wan
E-mail	msawan@cityu.edu.hk

Notice: Accepted version subject to English editing.

Substituting (4) into (3) leads to the following estimating equation for $\boldsymbol{\beta}$ in model (1):

$$\sum_{i=1}^{n} \delta_i I(J_i = 2) \left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} \boldsymbol{Z}_j I(\log \tilde{T}_j - \boldsymbol{Z}_j^T \boldsymbol{\beta} \geq \log \tilde{T}_i - \boldsymbol{Z}_i^T \boldsymbol{\beta})}{\sum_{j=1}^{n} I(\log \tilde{T}_j - \boldsymbol{Z}_j^T \boldsymbol{\beta} \geq \log \tilde{T}_i - \boldsymbol{Z}_i^T \boldsymbol{\beta})} \right] = 0. \tag{5}$$

Note that the l.h.s. of (5) is not monotone in $\boldsymbol{\beta}$, and this may produce multiple solutions of $\boldsymbol{\beta}$. To reconcile this difficulty, we consider the following monotone rank estimating equation analogous to that proposed by Fygenson and Ritov (1994) for censored data:

$$\tilde{\boldsymbol{U}}_n(\boldsymbol{\beta}) \equiv n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j) I\big\{ \log(\tilde{T}_j) - \boldsymbol{Z}_j^T \boldsymbol{\beta} \geq \log(\tilde{T}_i) - \boldsymbol{Z}_i^T \boldsymbol{\beta} \big\} = 0. \tag{6}$$

Although the l.h.s. of (6) is monotone in $\boldsymbol{\beta}$, it is still discontinuous with respect to $\boldsymbol{\beta}$ due to the presence of an indicator (jump) function in it. A range of well-developed algorithms including the brutal search method, Nelder-Mead method and linear programming method developed by Jin, Lin, Wei and Ying (2003) can be used for computing $\widehat{\boldsymbol{\beta}}$. However, as the asymptotic covariance matrix of the estimators involves the hazards function of an unspecified error distribution, direct estimation of the covariance matrix requires an estimate of this hazards function. Recognising that this estimate can be highly unstable, Jin, Lin, Wei and Ying (2003) proposed a resampling method to estimate the covariance matrix that eliminates the estimation of the hazards function but the computation efforts involved for the resampling method can be immense, especially with large data-sets. This motivates us to develop a differentiable estimating equation to approximate (6). Specifically, define $r_i^{\boldsymbol{\beta}} = \log(\tilde{T}_i) - \boldsymbol{Z}_i^T \boldsymbol{\beta}, i = 1, 2, \ldots, n$, along the lines of Heller (2007), we consider an approximation to the indicator function $I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})$ by a local distribution function $S((r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}})/\sigma_n)$, where $S(u)$ is non-decreasing and satisfies the conditions $\lim_{u \to \infty} S(u) = 1$ and $\lim_{u \to -\infty} S(u) = 0$, where σ_n is a sequence of strictly positive and decreasing numbers satisfying $\lim_{n \to \infty} \sigma_n = 0$. Clearly, when $r_j^{\boldsymbol{\beta}} > r_i^{\boldsymbol{\beta}}$, $S((r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}})/\sigma_n) \to 1$ as $n \to \infty$; on the other hand, when $r_j^{\boldsymbol{\beta}} < r_i^{\boldsymbol{\beta}}$, $S((r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}})/\sigma_n) \to 0$ as $n \to \infty$. A smoothed version of (6) is thus given by

$$n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j) S\Big(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n} \Big) = 0. \tag{7}$$

3 Methods for Handling Missing Causes of Failure

When the causes of failure are only partially available, the estimating equation (7) cannot be applied because J_i's are not observed for all i's. Now, let R_i be the complete-case indicator that is equal to 1 when either $\delta_i = 0$, or $\delta_i = 1$ and J_i is observed, and equal to 0 otherwise. Thus, when the causes of failure are not completely observed, the right-censored competing risks data set comprises i.i.d. observations of $\{(\tilde{T}_i, \delta_i, \boldsymbol{Z}_i, A_i, R_i, R_i \delta_i J_i), i = 1, \ldots, n\}$, where A_i's are some auxiliary covariates that may be useful for predicting the missing failure type.

We assume that the cause of failure is MAR (Rubin, 1976). That is, given $\delta_i = 1$ and $\boldsymbol{W}_i = (\tilde{T}_i, \boldsymbol{Z}_i^T, A_i)^T$, the probability that the failure cause of the i^{th} subject is missing depends only on the observed $\boldsymbol{W}_i$, but not on the unobserved J_i. Specifically, we assume that the failure cause missing probability is given by

$$r(\boldsymbol{W}_i) = P(R_i = 1 | J_i, \delta_i = 1, \boldsymbol{W}_i) = P(R_i = 1 | \delta_i = 1, \boldsymbol{W}_i). \tag{8}$$

Although the MAR assumption is more restrictive than nonignorable missingness, MAR is justified in many practical situations, and there is a large collection of literature that uses the MAR assumption as the baseline for analysis. Recent examples include Aerts, Claeskens, Hens and Molenberghs (2002), Wang and Rao (2002), Chen, Ibrahim and Shao (2004), Qi, Wang and Prentice (2005), Lu and Copas (2005), Zhou, Wan and Wang (2008), among others. In the remainder of this section, we develop three methods for dealing with missing data in the context of competing risks data.

3.1 Inverse probability weighting

Write $\boldsymbol{Q}_i = (\boldsymbol{W}_i^T, \delta_i)^T$. From Horvitz and Thompson (1952), note that

$$M_i^{(1)}(t) \equiv \frac{R_i}{\pi(\boldsymbol{Q}_i)} N_i(t) - \int_{-\infty}^{t} Y_i(u)\lambda(u)du, \; i = 1, 2, \ldots, n,$$

are mean zero processes, where $\pi(\boldsymbol{Q}_i) = P(R_i = 1|\delta_i, \boldsymbol{W}_i) = \delta_i r(\boldsymbol{W}_i) + (1 - \delta_i)$. By derivations similar to those in Section 2, this leads to the following inverse probability weighted (IPW) estimating equation for $\boldsymbol{\beta}$:

$$n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{R_i}{\pi(\boldsymbol{Q}_i)} \delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j) S\left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n}\right) = 0. \tag{9}$$

In practice, $r(\boldsymbol{W}_i)$ is often unknown. We may estimate $r(\boldsymbol{W}_i)$ parametrically as in Gao and Tsiatis (2005), Lu and Liang (2008) and Sun, Wang and Gilbert (2012), or non-parametrically as in Qi, Wang and Prentice (2005), Zhou, Wan and Wang (2008), and Song, Sun, Mu and Dinse (2010). Here, we adopt the non-parametric approach which has the advantage over its parametric counterpart of being less prone to biases arising from model mis-specification. We use a kernel method and assume that d is the size of the continuous elements in $\boldsymbol{W}_i$ and $k(u)$ is a rth-order ($r > d$) kernel function with compact support that satisfies the following conditions: $\int k(u)du = 1$, $\int u^m k(u)du = 0$ for $m = 1, 2, \ldots, r-1$, $\int u^r k(u)du \neq 0$ and $\int k^2(u)du < \infty$. As well, for any $\boldsymbol{u} = (u_1, u_2, \ldots, u_d) \in R^d$, define $K_h(\boldsymbol{u}) = \frac{1}{h^d} \prod_{i=1}^{d} k(u_i/h)$, where h is a bandwidth sequence that satisties $nh^{2r} \to 0$ and $nh^{2d} \to \infty$ as $n \to \infty$. The Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964) of $r(\boldsymbol{w})$ is then given by

$$\widehat{r}(\boldsymbol{w}) = \widehat{G}_n^{-1}(\boldsymbol{w}) \frac{1}{n} \sum_{i=1}^{n} R_i \delta_i K_h(\boldsymbol{w}_1 - \boldsymbol{W}_{1i}) I(\boldsymbol{W}_{2i} = \boldsymbol{w}_2), \tag{10}$$

where $\widehat{G}_n(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n} \delta_i K_h(\boldsymbol{w}_1 - \boldsymbol{W}_{1i}) I(\boldsymbol{W}_{2i} = \boldsymbol{w}_2)$, $\boldsymbol{w} = (\boldsymbol{w}_1, \boldsymbol{w}_2)$, and $\boldsymbol{W}_{1i}$ and $\boldsymbol{W}_{2i}$ are matrices that contain the continuous and discrete elements of $\boldsymbol{W}_i$ respectively. Substituting the estimator $\widehat{r}(\boldsymbol{W}_i)$ into (9) leads to the following IPW estimating equation for $\boldsymbol{\beta}$:

$$\boldsymbol{U}_1(\boldsymbol{\beta}) \equiv n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{R_i}{\widehat{\pi}(\boldsymbol{Q}_i)} \delta_i I(J_i = 2)(\boldsymbol{Z}_i - \boldsymbol{Z}_j) S\left(\frac{r_j^{\boldsymbol{\beta}} - r_i^{\boldsymbol{\beta}}}{\sigma_n}\right) = 0, \tag{11}$$

where $\widehat{\pi}(\boldsymbol{Q}_i) = \delta_i \widehat{r}(\boldsymbol{W}_i) + (1 - \delta_i)$.

Denote the solution of (11) as $\widehat{\boldsymbol{\beta}}_{IPW}$. The development of an asymptotic theory for $\widehat{\boldsymbol{\beta}}_{IPW}$ (as well as that of the other estimators in the subsequent sections) requires the following conditions:

(C1) The covariate vector, $\boldsymbol{Z}_1$, is bounded, and there exists a constant M such that, $\| E(\boldsymbol{Z}_1 - \boldsymbol{Z}_2)(\boldsymbol{Z}_1 - \boldsymbol{Z}_2)^T \| < M < \infty$, and the parameter $\boldsymbol{\beta}$ lies in a compact set $\mathcal{B}$.