Incorporating Follow-up Time In M-Estimation For Survival Data

Glenn Heller

Department of Epidemiology and Biostatistics, Box 44, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, U.S.A.

Telephone number: 646-735-8112

Fax number: 646-735-0010

email address: heller@biosta.mskcc.org

Incorporating Follow-up Time In M-Estimation For Survival Data

Abstract

It has been approximately thirty years since D.R. Cox introduced the proportional hazards method to model the relationship between covariates and survival time. However, the proportional hazards model has limited value when the proportionality assumption is violated. Over the years, there have many been many alternative proposals to the proportional hazards regression model for the case of right censored survival data, but to date none have demonstrated widespread acceptance. In general, problems encountered in these methods include their computational algorithms or evaluation of their asymptotic properties. In this work, an estimating equation based on a U-statistic of degree 2 is proposed. It is easy to implement and the U-statistic framework provides a straightforward development of asymptotic inferential theory for the regression parameters.

Keywords: administrative censoring, estimating equation, kernel smoothing, U-statistic

1. Introduction

It has been approximately thirty years since D.R. Cox introduced the proportional hazards method to model the relationship between covariates and survival time (Cox 1972). It remains to this day the predominant model applied to survival data. The widespread popularity of the proportional hazards methodology stems from the interpretation of the regression coefficient as a relative risk parameter constant with respect to time, the development of asymptotic inferential procedures that are easy to implement with the advent of readily available software, and the efficiency of the regression parameter estimates for a wide range of underlying hazard functions. Additionally, no parametric assumption of the underlying hazard function is required. However, the proportional hazards model has limited value when the proportionality assumption is violated. Although there are many techniques developed to determine if the proportionality assumption holds, there is no methodology such as Box-Cox (1964) to transform the data to the proportional hazards model in order to bring it into line with this assumption. Thus, increasing the scope of alternative regression models when the proportionality assumption does not hold will benefit analyses of survival data.

An alternative semiparametric approach is the accelerated failure time model

$$\log t_i = \boldsymbol{\beta}_0^T \mathbf{x}_i + \epsilon_i \qquad i = 1, \dots, n \tag{1}$$

where the $\{\epsilon_i\}$ are independent identically distributed errors with distribution function F, and $\mathbf{x}_i$ is independent of ϵ_i. The model is termed semiparametric because the functional form of the response surface is assumed known, while the error distribution F is unknown.

A natural approach for inference on $\boldsymbol{\beta}_0$ with F unknown is the estimating equation. In the accelerated failure time regression model with right censored data, the estimating equation approach has been used for L-estimation (Ying et al. 1995), R-estimation (Tsiatis 1990, Ying 1993, Fygenson and Ritov 1994) and M-estimation (Buckley and James 1979, Ritov 1990, Lai and Ying 1992). A baseline requirement of the estimating function, needed for the development of asymptotic inferential theory, is that it have asymptotic mean zero. The L or quantile based estimating equation of Ying et al. (1995) satisfies this requirement by incorporating a consistent estimate for the distribution function of random censoring times. The M-estimate, which uses 'real time' data, uses the Kaplan-Meier estimate of the stochastic error term. R estimation inverts the rank based test statistics to produce an asymptotically unbiased estimating equation. The consistency and asymptotic normality of L, M, and R estimators in the presence of right censored data have been developed in the cited works above, and provide the large sample foundation for inference of the regression parameter $\boldsymbol{\beta}_0$.

The practical application of these techniques has been hindered due to complications stemming from the discontinuous nature of the estimating functions and estimation of the asymptotic variance of the regression estimate; both problems are attributable to censoring. In L estimation, Ying et al. (1995) minimize the L_2 norm of the estimating function to estimate $\boldsymbol{\beta}_0$. However, they are cautious as to the reliability of this search due to discontinuities in the estimating function. For R and M estimation, computation of the estimate $\hat{\boldsymbol{\beta}}$, is typically based on a Newton-Raphson type algorithm that searches for a zero solution, or zero crossing, of the estimating function. Newton-Raphson is based on a Taylor series expansion, and requires computation of the first derivative of the estimating function with respect to $\boldsymbol{\beta}$. For R estimates, application of this algorithm is complicated by the fact that the estimating functions are step functions in $\boldsymbol{\beta}$. Tsiatis (1990) demonstrates that the step function may be asymptotically approximated by a linear function in a neighborhood of $\boldsymbol{\beta}_0$. The resulting regression estimates involve kernel smoothing of the hazard function of the error term (ϵ), which is unstable in the right tail when censoring is present (Hess et al. 1999). Gray (2000) uses a penalized likelihood for the hazard estimation. In both cases, the consistency of the asymptotic variance estimate, which is a function of the hazard, has not been established. Fygenson and Ritov (1994) propose a family of rank based estimating functions that are monotone in $\boldsymbol{\beta}$. They replace the smoothing employed by Tsiatis with numerical differentiation for computing the Taylor expansion. Although estimation of the regression coefficient may alternatively be accomplished through a minimization algorithm, the asymptotic variance estimate of $\hat{\boldsymbol{\beta}}$ is complicated by the use of numerical differentiation, particularly in the multiple covariate case.

The estimating equation most easily adapted to the Newton-Raphson zero finding algorithm is the M estimating equation, which uses real time data. In the presence of right censored data, these equations incorporate the Kaplan-Meier estimate of the regression residual $\epsilon = t - \boldsymbol{\beta}^T \mathbf{x}$. In addition to the discontinuities produced by the Kaplan-Meier estimate, its introduction into the estimating equation results in an asymptotic variance of the regression estimate that contains the density and the density derivative of ϵ. Again, due to censoring, estimation of these functions are unstable in the right tail.

In this paper, an unbiased M estimating equation is created using the failure times and adminstrative or lost to follow up censoring times. Estimation of a density or a density derivative function is not required since it does not use the Kaplan-Meier estimate in the estimating function. The additional censoring information, which is not related to the parameter of interest, reduces the efficiency of the proposed estimator. However, the development of an asymptotic inferential theory, including estimation of the asymptotic variance, is straightforward and the estimating procedure is easy to implement. In section 2 of the paper, administrative censoring is introduced. In section 3, the M estimating equation is presented and the asymptotic distribution of the regression estimate is developed. In section 4, model based quantile prediction of survival time is proposed. In section 5, simulations are performed to examine the finite sample adequacy of the parameter estimate and coverage based on asymptotic confidence intervals. Section 6 includes concluding remarks.

2. Censoring Times

In the conventional approach to survival analysis, each individual is associated with a bivariate stochastic vector of survival and censoring times $\{t_i, c_i\}_{i=1}^n$. The minimum time and censoring indicator are observed for each subject

$$y_i = \min(t_i, c_i) \qquad \delta_i = I(t_i \leq c_i) \qquad i = 1, \ldots, n.$$

The censoring time may be considered the result of the operation $c = \min(c^{[1]}, c^{[2]})$, where $c^{[1]}$ is defined as the time from study entry to study closure, often termed administrative censoring, and $c^{[2]}$ as the time from study entry to a noninformative lost to follow up (Miller 1981). Thus the censoring time c is independent of survival time.

In this paper, censoring times are explicitly used to create a mean zero estimating function. For each subject, a censoring time c is determined. For subjects who have not failed $(\delta = 0)$, $c = \min(c^{[1]}, c^{[2]})$. This coincides with the conventional use of independent censoring in survival analysis. However, subjects who have failed $(\delta = 1)$ are assigned the administrative censoring time $c = c^{[1]}$. Under this construction, the censoring time c remains independent of the survival time t. The use of administrative censoring in survival analysis has been proposed previously in works including: Robins and Tsiatis (1991), Robins (1992), Oakes (1993), and Joffe (2001).

3. Estimation and Inference for Regression Parameter $\boldsymbol{\beta}_0$

We begin by estimating the regression parameter $\boldsymbol{\beta}_0$ from the accelerated failure time model in equation (1) for the uncensored data case. Although it is unnecessary to present a new estimating equation in this situation, the equation and the attendant inference procedure are presented to provide the framework and intuition of the methodology for the case of right censored data.

3.1 No Censoring

To estimate $\boldsymbol{\beta}_0$ in the case of no censoring, consider the estimating function

$$S_n^*(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j (\mathbf{x}_i - \mathbf{x}_j)(\epsilon_i^{\boldsymbol{\beta}} - \epsilon_j^{\boldsymbol{\beta}}) \tag{2}$$

where $\epsilon^{\boldsymbol{\beta}} = \log t - \boldsymbol{\beta}^T \mathbf{x}$ is termed the regression residual. At $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, the estimating function has expectation zero, $S_n^*(\boldsymbol{\beta})$ is continuous in $\boldsymbol{\beta}$, and $S_n^*(\boldsymbol{\beta}_0)$ is a U-statistic of degree 2 (Bickel et al. 1986). In addition, the derivative of the function with respect to $\boldsymbol{\beta}$, $\partial S_n^*(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = -n^{-3/2} \sum_i \sum_j (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$ is negative definite, demonstrating that S_n^* is a monotone function of $\boldsymbol{\beta}$. Zero mean, continuity, and monotonicity provide the foundation for the asymptotic theory of M-estimation, and insure that the neighborhood of $\boldsymbol{\beta}_0$ is located in an algorithmic search for the zero solution to the estimating equation. The M-estimate $\hat{\boldsymbol{\beta}}$ is the solution to the estimating equation $S_n^*(\boldsymbol{\beta}) = 0$.

Employment of the asymptotic distribution theory for the U-statistic $S_n^*(\boldsymbol{\beta}_0)$ in conjunction with a Taylor series expansion of $S_n^*(\hat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}_0$ demonstrates that $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ has a limiting normal distribution with mean zero. From a numerical viewpoint, the monotonicity of the estimating function insures that the solution to the estimating equation is unique and that the Newton-Raphson algorithm converges at a quadratic rate.

3.2 Right Censoring

To present the estimating function in the presence of right censored data, we introduce the censoring residual $m^{\boldsymbol{\beta}} = \log c - \boldsymbol{\beta}^T \mathbf{x}$. In this setup, $m^{\boldsymbol{\beta}}$ is observable up to the unknown parameter $\boldsymbol{\beta}$. Since some failure times are censored, not all regression residuals are observable, requiring an alternative estimating equation than the one presented in (2).

Suppose the estimating function is restricted to regression residuals corresponding to observed failure times. This corresponds to the set $\{i : \epsilon_i^{\boldsymbol{\beta}} < m_i^{\boldsymbol{\beta}}\}$. One such estimating function is

$$\sum_i \sum_j \delta_i \delta_j (\mathbf{x}_i - \mathbf{x}_j)(\epsilon_i^{\boldsymbol{\beta}} - \epsilon_j^{\boldsymbol{\beta}}).$$

This restriction to failure time residuals produces a biased estimating equation, with the bias a result of the unequal censoring residuals. To create equal restriction intervals on the regression residuals, the following estimating equation is presented

$$\tilde{S}_n(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i (\mathbf{x}_i - \mathbf{x}_j)(\epsilon_i^{\boldsymbol{\beta}} - \epsilon_j^{\boldsymbol{\beta}}) I(\epsilon_j^{\boldsymbol{\beta}} < m_i^{\boldsymbol{\beta}}) I(m_j^{\boldsymbol{\beta}} \geq m_i^{\boldsymbol{\beta}}). \qquad (3)$$

The indicator functions $I(\epsilon_j^{\boldsymbol{\beta}} < m_i^{\boldsymbol{\beta}})I(m_j^{\boldsymbol{\beta}} \geq m_i^{\boldsymbol{\beta}})$ are used to truncate the regression residual $\epsilon_j^{\boldsymbol{\beta}}$ at $m_i^{\boldsymbol{\beta}}$ and to select $\delta_j = 1$. This matches the truncation of $\epsilon_i^{\boldsymbol{\beta}}$ at $m_i^{\boldsymbol{\beta}}$, produced by the selection indicator $\delta_i = 1$. Thus, the regression residuals $(\epsilon_i^{\boldsymbol{\beta}}, \epsilon_j^{\boldsymbol{\beta}})$ corresponding to observed failures are restricted to the same interval with upper bound $m_i^{\boldsymbol{\beta}}$. These equal restriction intervals produce a mean zero estimating function at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. As a result, the following lemma is presented.

Lemma: The estimating function $\tilde{S}_n(\boldsymbol{\beta}_0)$ converges to a normal mean zero random variable. In the appendix it is shown that $E[\tilde{S}_n(\boldsymbol{\beta}_0)] = 0$. Since $\tilde{S}_n(\boldsymbol{\beta}_0)$ is a U-statistic of degree 2, the lemma follows directly from the asymptotic distribution theory of U-statistics.

The estimating function $\tilde{S}_n(\boldsymbol{\beta})$ is piecewise linear in $\boldsymbol{\beta}$. It has jumps at $2n^2p$ points in the domain $\boldsymbol{\mathcal{B}}$ of $\boldsymbol{\beta}$; for all other $\boldsymbol{\beta} \in \bar{\boldsymbol{\mathcal{B}}} \subset \boldsymbol{\mathcal{B}}$, the estimating function is differentiable. Since the number of jumps is countable, it is assumed that $\boldsymbol{\beta}_0 \in \bar{\boldsymbol{\mathcal{B}}}$. Thus from an asymptotic viewpoint, the distributional properties of the estimated $\boldsymbol{\beta}_0$ are straightforward. However, there are practical difficulties with this unsmooth estimating function. Although the Newton-Raphson algorithm can be applied to find the zero solution to $\tilde{S}_n(\boldsymbol{\beta})$, its piecewise linearity will effect the algorithm's iterative search and computation of the asymptotic variance of the estimate. Both are a function of the piecewise constant Hessian matrix

$$\tilde{A}_n(\boldsymbol{\beta}) = -n^{-3/2} \sum_i \sum_j \delta_i (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T I(\epsilon_j^{\boldsymbol{\beta}} < m_i^{\boldsymbol{\beta}}) I(m_j^{\boldsymbol{\beta}} \geq m_i^{\boldsymbol{\beta}}) \qquad \boldsymbol{\beta} \in \bar{\mathcal{B}}.$$

As a result of the potential problems caused by the discontinuities, a smoothed analog of $\tilde{S}_n(\boldsymbol{\beta})$ is proposed. To ease the notation for the smoothed estimating equation, all pairwise differences $w_i - w_j$ are written as w_{ij}. Then the smoothed estimating equation is

$$S_n(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i \delta_j \mathbf{x}_{ij} \epsilon_{ij}^{\boldsymbol{\beta}} \{G_{\boldsymbol{\beta}^T \mathbf{x}_{ji}, h^2}(c_{ji}) - G_{\boldsymbol{\beta}^T \mathbf{x}_{ji}, h^2}(y_j - c_i)\} \tag{4}$$

where $G_{\boldsymbol{\beta}^T \mathbf{x}_{ji}, h^2}$ is a local distribution function from a location-scale family, symmetric about $\boldsymbol{\beta}^T \mathbf{x}_{ji}$ with scale parameter h^2. Using smoothing terminology, h is the bandwidth. The piecewise linear function $\tilde{S}_n(\boldsymbol{\beta})$ and the smoothed function $S_n(\boldsymbol{\beta})$ are asymptotically equivalent, assuming the bandwidth h goes to zero at a sufficiently fast rate. Heuristically, this is demonstrated by rewriting $\tilde{S}_n(\boldsymbol{\beta})$ as

$$\tilde{S}_n(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i \delta_j \mathbf{x}_{ij} (\epsilon_i^{\boldsymbol{\beta}} - \epsilon_j^{\boldsymbol{\beta}}) I\{(y_j - c_i) < \boldsymbol{\beta}^T \mathbf{x}_{ji} \leq c_{ji}\}$$

and noting that as h goes to zero in equation (4), the local distribution function G converges to a step function with a single jump at $\boldsymbol{\beta}^T \mathbf{x}_{ji}$, and therefore $G_{\boldsymbol{\beta}^T \mathbf{x}_{ji}, h^2}(c_{ji}) - G_{\boldsymbol{\beta}^T \mathbf{x}_{ji}, h^2}(y_j - c_i)$ converges to $I\{(y_j - c_i) < \boldsymbol{\beta}^T \mathbf{x}_{ji} \leq c_{ji}\}$. Further details on the rate of convergence are provided in the appendix.

We define the estimate $\hat{\boldsymbol{\beta}}$, as the solution to the estimating equation $S_n(\boldsymbol{\beta}) = 0$. The following theorem summarizes the asymptotic distribution of the solution to the estimating equation $S_n(\boldsymbol{\beta}) = 0$. A proof of the theorem is provided in the appendix.

Theorem: If $\hat{\boldsymbol{\beta}}$ lies in a compact neighborhood of $\boldsymbol{\beta}_0$, then under the conditions stated in the appendix, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}VA^{-1})$, where $A = \mathrm{E}\{n^{-1/2} \partial S_n(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}\}|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0}$ and $V = n^{-1} \mathrm{var}\{S_n(\boldsymbol{\beta}_0)\}$.

Since the estimating equation is not monotone in $\boldsymbol{\beta}$, the asymptotic distribution theory requires a consistent estimate of $\boldsymbol{\beta}_0$ to start the Newton-Raphson algorithm. Based on the results in Fygenson and Ritov (1994), an initial consistent estimate $\hat{\boldsymbol{\beta}}_I$ may be determined by choosing $\boldsymbol{\beta}$ to minimize $\parallel \sum_i \sum_j \delta_i \boldsymbol{\beta}^T \mathbf{x}_{ij} I\{\boldsymbol{\beta}^T \mathbf{x}_{ij} > y_{ij}\} \parallel$. The Nelder-Mead simplex algorithm (1965) is used to minimize this convex function. As noted earlier, although estimation of the regression coefficient is straightforward, the use of numerical differentiation to compute the asymptotic variance makes this approach less useful, particularly in the multiple covariate case.

The smoothing in the estimating equation (4) is performed on the difference in linear functionals $\boldsymbol{\beta}^T \mathbf{x}_{ji} = \boldsymbol{\beta}^T \mathbf{x}_j - \boldsymbol{\beta}^T \mathbf{x}_i$, a one-dimensional variable. A Gaussian local distribution function is used with bandwidth $h = \hat{\sigma}^2 n^{-1/3}$, where $\hat{\sigma}^2 = \sum_i \sum_j (\hat{\boldsymbol{\beta}}_I^T \mathbf{x}_{ji})^2 / n^2$. Note that although there may be multiple covariates in the model, the curse of dimensionality does not play a role since the smoothing is performed on the linear functional $\boldsymbol{\beta}^T \mathbf{x}_{ji}$. This use of smoothing is distinguished from the smoothing needed to estimate the residual density function in prior R and M estimating equation approaches alluded to earlier. In those situations, estimation of the entire error density function was problematic due to censoring. In the present situation, the smoothing operation is not on the residual or any function of survival time, and hence, the ensuing problems do not apply. This use of smoothing has also been proposed in the context of single-index regression models (Hardle et al. 1993).

4. Prediction

The accelerated failure time model enables the user to predict characteristics of the survival time distribution conditional on covariates. For example, the conditional median log survival time is represented by

$$\mathrm{md}(\log t|\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{x} + \mathrm{md}(\epsilon|\mathbf{x}).$$

Since $\mathbf{x}$ is independent of ϵ, one can use Kaplan-Meier to estimate $\text{md}(\log t - \hat{\boldsymbol{\beta}}^T\mathbf{x})$, which is denoted by $\hat{\alpha}_{.50}$. It follows that the conditional median log survival time may be estimated by

$$\hat{\text{md}}(\log t|\mathbf{x}) = \hat{\boldsymbol{\beta}}^T\mathbf{x} + \hat{\alpha}_{.50},$$

where $\hat{\boldsymbol{\beta}}$ was estimated in the previous section. Any quantile can be used for prediction, as long as it is well estimated by Kaplan-Meier. Thus, the amount of censoring will dictate how far out in the survival distribution accurate prediction of the conditional quantiles can be performed. If over 50% of a data set contains censored survival times, a quantile from the survival distribution larger than .50 should be used for prediction in the accelerated failure time model.

5. Prostate Cancer Example

A total of 363 subjects with prostate cancer that metastasized to the bone were treated at Memorial Sloan-Kettering Cancer Center between the years 1989 and 2000. To date, 324 patients have died. An important prognostic factor in this patient population is the enzyme alkaline phosphatase. Alkaline phosphatase is found in both the bone and the liver, and for this patient population, high levels are indicative of an increased tumor burden located in the bone. The relationship between log alkaline phosphatase and log survival time is explored using the conditional median response accelerated failure time model.

To determine the covariate specification in the accelerated failure time model, a kernel smoothed Kaplan-Meier estimate of the conditional median survival time was computed. A graphic depiction of this relationship is presented as a dashed line in Figure 1. The curve was generated using a Gaussian kernel with bandwidth chosen equal to $2\hat{\sigma}n^{-1/5}$, where $\hat{\sigma}$ is the estimated standard deviation of log alkaline phosphatase. The figure suggests that for log alkaline phosphatase values in the range 3-4, its relationship with survival time is constant, and the relationship is concave thereafter. As a result, we considered the specification $\text{md}(t) = \alpha \exp\{-\beta(x - 3.5)^2\}$ or in the accelerated failure time model form, $\log t = -\beta(x - 3.5)^2 + \epsilon$, where x represents log alkaline phosphatase and the $\text{md}\{\exp(\epsilon)\} = \alpha$. The solid curve in Figure 1, represents the fit from the accelerated failure time model, and demonstrates that this model provides a good fit to the data. The estimated coefficient was $\hat{\beta} = -0.067$ and the estimated $\text{se}(\hat{\beta}) = 0.002$, confirming the anticipated strong relationship between log alkaline phosphatase and survival time in this patient population. For the purposes of prediction, the intercept coefficient estimate was $\hat{\alpha} = 18.92$.

6. Simulations

A simulation study was performed to assess the finite sample performance of the estimate $\hat{\beta}$ and the accuracy of the asymptotic normal approximation developed in the theorem. The log survival times were generated from a single covariate linear regression model $\log t_i = \beta_0 x_i + \epsilon_i$; the regression parameter β_0 was set equal to 2.0. The ϵ_i were independent identically distributed as either normal or extreme value random variables with mean 1 and variance σ^2. The strength of the regression was varied from strong to weak by choosing σ to be $\{0.25, 0.50, 1.0, 2.0, 4.0\}$. The censoring times were determined by first generating values from a uniform distribution $(0, \tau)$, and then taking the log of these values. The choice of τ determines the percentage of censored observations in each replication. The upper terminal (τ) of the uniform was chosen to produce censoring proportions of $\{0.0, 0.25, 0.50, 0.75\}$. For all simulations, the sample size was $n = 200$, with x taking values -7.96(.08)7.96. There were 5000 replications for each simulation.

The results of the simulations presented in Tables 1a and 1b demonstrate that the accuracy of the estimator and the asymptotic normal approximation are good. The bias of the estimator is small, except in the case that the censoring proportion and the dispersion parameter are high. This observation, which was made earlier in the general case of M-estimators with survival data (Heller and Simonoff 1990), is attributable to the asymmetric effect of large σ on right censored data. Large positive ϵ produce censored survival times and large negative ϵ produce observed survival times. Tables 1a and 1b also show good concordance between the estimated standard error ($\hat{\Sigma}$) and the simulation standard error, and that the empirical coverage probabilities of the asymptotic 95% confidence intervals are close to the nominal .95 level.

TABLE 1A The columns in the table represent: the strength of the regression relationship, the percent censored, the bias of $\hat{\beta}$, the standard deviation of the simulation estimates of $\hat{\beta}$, the average estimated standard error, and the empirical coverage probability.

Normal error

σ	% censor	$\mathrm{E}(\hat{\beta} - \beta_0)$	$V^{1/2}(\hat{\beta})$	$\mathrm{E}(\hat{\Sigma}^{1/2})$	Coverage
0.25					
	0	.000	.000	.000	.944
	25	.000	.000	.000	.940
	50	.000	.010	.010	.944
	75	-.001	.036	.035	.936
0.50					
	0	.000	.010	.010	.948
	25	-.001	.014	.014	.943
	50	-.001	.022	.020	.942
	75	-.001	.072	.069	.923
1.00					
	0	.001	.014	.014	.950
	25	-.001	.026	.024	.932
	50	-.003	.042	.041	.937
	75	-.015	.134	.134	.928
2.00					
	0	.003	.032	.030	.938
	25	-.001	.052	.049	.930
	50	.004	.083	.081	.944
	75	.003	.218	.252	.959
4.00					
	0	.005	.065	.061	.936
	25	.015	.099	.095	.936
	50	.037	.147	.155	.948
	75	.085	.370	.419	.963

Table 1b

σ	% censor	$E(\hat{\beta} - \beta_0)$	$V^{1/2}(\hat{\boldsymbol{\beta}})$	$E(\hat{\Sigma}^{1/2})$	Coverage
		Extreme value error			
0.25					
	0	.000	.000	.000	.946
	25	.000	.000	.000	.945
	50	.000	.010	.010	.939
	75	-.001	.035	.035	.936
0.50					
	0	.000	.010	.010	.945
	25	.000	.014	.014	.944
	50	-.001	.020	.020	.942
	75	-.008	.069	.068	.932
1.00					
	0	.000	.014	.014	.947
	25	-.001	.024	.024	.943
	50	-.002	.041	.040	.939
	75	-.009	.128	.130	.933
2.00					
	0	.000	.032	.030	.947
	25	.000	.049	.047	.935
	50	.003	.079	.078	.936
	75	.031	.220	.241	.952
4.00					
	0	.002	.062	.061	.945
	25	.012	.097	.091	.927
	50	.037	.149	.147	.940
	75	.197	1.14	.496	.936

7. Discussion

The proposed method enables estimation of the intercept term and regression coefficients using M-estimation with real time survival data. The estimation and inference are based on the set of assumptions that the errors $\{\epsilon\}$ are independent and identically distributed, and the mean response function is correctly specified. In comparison, the proportional hazards model requires an independent identically distributed error term, a correct specification of the conditional hazard function, along with the proportionality assumption. The proportional hazards assumption, which can be stated as: there exists a monotone transformation h such that, $h(t) = \boldsymbol{\beta}^T\mathbf{x} + \epsilon$, where ϵ is a standard extreme value random variable, enables estimation and inference within a likelihood framework. As a result, an efficient estimate of the regression coefficient $\boldsymbol{\beta}$ can be produced.

It is anticipated that the reduced number of assumptions will translate into a wider range of applications for the M-estimate relative to the proportional hazards estimate. However, if the proportional hazards assumption is appropriate, its application will provide a more efficient estimate. This contrast follows the standard tradeoff of efficiency versus robustness when deciding between likelihood and estimating equation methods. Future research will examine the relative efficiency of the methods under various forms of model misspecification. Additional consideration will be given to the choice of a weight function in the estimating equation. Weights could be chosen to produce a more efficient estimate or to produce a negative definite Hessian $A_n(\boldsymbol{\beta}) = \partial S_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$. Finally, the relationship between bandwidth selection and efficiency will be explored by choosing h to minimize the asymptotic variance of $\hat{\boldsymbol{\beta}}$ subject to the constraint that nh^4 converges to zero.

Appendix

Lemma: $E[\tilde{S}_n(\boldsymbol{\beta}_0)] = 0$

$$n^{3/2}\tilde{S}_n(\boldsymbol{\beta}_0) = \sum_i \sum_j \mathbf{x}_{ij}\{\delta_i \epsilon_i I(\epsilon_j < m_i) - \delta_i \epsilon_j I(\epsilon_j < m_i)\} I(m_j \geq m_i)$$

Taking the expectation conditional on $(c_i, c_j, \mathbf{x}_i, \mathbf{x}_j)$

$$= \sum_i \sum_j \mathbf{x}_{ij}\{F^2(m_i)E(\epsilon_i|\epsilon_i < m_i) - F^2(m_i)E(\epsilon_j|\epsilon_j < m_i)\} I(m_j \geq m_i) = 0.$$

Theorem: $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(A^{-1}VA^{-1})$.

The proof is divided into three parts: a) the asymptotic equivalence between the smooth and unsmooth estimating functions; b) the consistency of the estimator; and c) the asymptotic distribution of the estimator.

For the proof below, let $G_{\boldsymbol{\beta}^T \mathbf{x}_{ji}, h^2}(z)$ be a local distribution function from a location-scale family, symmetric about $\boldsymbol{\beta}^T \mathbf{x}_{ji}$ with scale (bandwidth) parameter h^2, $\Phi(z) = G_{0,1}(z)$, and $\phi(z) = \partial \Phi(z)/\partial z$.

The following conditions are required for the proof.

(i) The random vector $(\mathbf{x}, t, c)$ lies in a $p + 2$ dimensional bounded rectangle.

(ii) The parameter vector $\boldsymbol{\beta}$ lies in a p dimensional bounded rectangle.

(iii) $n^{-1/2}S_n(\boldsymbol{\beta})$ has a bounded first derivative $n^{-1/2}A_n(\boldsymbol{\beta})$ in a compact neighborhood of $\boldsymbol{\beta}_0$ - $\mathcal{N}(\boldsymbol{\beta}_0)$, with $n^{-1/2}A_n(\boldsymbol{\beta})$ nonzero in $\mathcal{N}(\boldsymbol{\beta}_0)$.

(iv) $\Phi(z)$ has bounded support and is Lipschitz continuous on its support. $\phi(z)$ is the kernel density, symmetric about zero, of order 2.

(v) As $n \to \infty$, the bandwidth h is chosen such that $nh^4 \to 0$.

a) $S_n(\boldsymbol{\beta}) = \tilde{S}_n(\boldsymbol{\beta}) + O_p(n^{1/2}h^2)$ uniformly in $\boldsymbol{\beta}$.

The smoothed estimating equation is

$$S_n(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i \delta_j \mathbf{x}_{ij} \epsilon_{ij} \{G_{\boldsymbol{\beta}^T \mathbf{x}_{ji}, h^2}(c_{ji}) - G_{\boldsymbol{\beta}^T \mathbf{x}_{ji}, h^2}(y_j - c_i)\}$$

where all pairwise differences $w_i - w_j$ are written as w_{ij}. The difference between the smoothed and unsmoothed estimating functions is written as

$$\begin{aligned} S_n(\boldsymbol{\beta}) - \tilde{S}_n(\boldsymbol{\beta}) &= n^{-3/2} \sum_i \sum_j \delta_i \delta_j \mathbf{x}_{ij} \epsilon_{ij} \left\{ \Phi\left(\frac{v_{ij} - 0}{h}\right) - I(0 < v_{ij}) \right\} \\ & -n^{-3/2} \sum_i \sum_j \delta_i \delta_j \mathbf{x}_{ij} \epsilon_{ij} \left\{ \Phi\left(\frac{u_{ij} - 0}{h}\right) - I(0 < u_{ij}) \right\} \end{aligned}$$

where $v_{ij} = c_{ji} - \boldsymbol{\beta}^T \mathbf{x}_{ji}$ and $u_{ij} = (y_j - c_i) - \boldsymbol{\beta}^T \mathbf{x}_{ji}$. To determine the order of magnitude of $S_n(\boldsymbol{\beta}) - \tilde{S}_n(\boldsymbol{\beta})$, the first component in this difference is examined; the same argument is applied to the second component. Using the Cauchy-Schwarz inequality and the bounding conditions,

$$n^{-3/2} \left\| \sum_i \sum_j \delta_i \delta_j \mathbf{x}_{ij} \epsilon_{ij} \left\{ \Phi\left(\frac{v_{ij} - 0}{h}\right) - I(0 < v_{ij}) \right\} \right\|$$

$$\leq M n^{-3/2} \left| \sum_i \sum_j \left\{ \Phi\left(\frac{v_{ij} - 0}{h}\right) - I(0 < v_{ij}) \right\} \right|$$

To simplify subsequent expressions, this is written as

$$\leq M n^{1/2} \left| \int_v \left\{ \Phi\left(\frac{v - 0}{h}\right) - I(0 < v) \right\} d\hat{F}_{n \times n}(v) \right| \tag{A.1}$$

where $\hat{F}_{n \times n}(v)$ is the empirical cumulative distribution function with jumps at each of the n^2 elements of v_{ij}. For $F(v) = \lim_{n \to \infty} \hat{F}_{n \times n}(v)$, we can further the inequality in (A.1) to

$$\begin{aligned} \leq \; & M n^{1/2} \left| \int_v \left\{ \Phi\left(\frac{v - 0}{h}\right) - I(0 < v) \right\} d[\hat{F}_{n \times n}(v) - F(v)] \right| \\ + \; & M n^{1/2} \left| \int_v \left\{ \Phi\left(\frac{v - 0}{h}\right) - I(0 < v) \right\} dF(v) \right| \end{aligned} \tag{A.2}$$

Let

$$U_1(h) = \int_v \Phi\left(\frac{v-0}{h}\right) d[\hat{F}_{n\times n}(v) - F(v)]$$

$$U_2(0) = \int_v I(0 < v) d[\hat{F}_{n\times n}(v) - F(v)]$$

$$B(h) = \int_v \Phi\left(\frac{v-0}{h}\right) dF(v) - [1 - F(0)]$$

so (A.2) is expressed as

$$\leq Mn^{1/2}|U_1(h) - U_2(0)| + Mn^{1/2}|B(h)|.$$

The order of magnitude of these terms is as follows.

For $U_1(h)$, a change of variable $z = h^{-1}v$ and integration by parts gives

$$U_1(h) = -\int_z \phi(z)[\hat{F}_{n\times n}(zh) - F(zh)]dz,$$

and hence

$$Mn^{1/2}|U_1(h) - U_2(0)| =$$
$$Mn^{1/2}\left|\int_z \phi(z)\left\{\left[\hat{F}_{n\times n}(zh) - F(zh)\right] - \left[\hat{F}_{n\times n}(0) - F(0)\right]\right\} dz\right|.$$

Using the results on oscillations of empirical processes (Shorack and Wellner, 1986; p.531)

$$Mn^{1/2}|U_1(h) - U_2(0)| = O_p\left(\left[h\log n\ \log\left(\frac{1}{h\log n}\right)\right]^{1/2}\right).$$

For $B(h)$, integration by parts and a two-term Taylor expansion around $h = 0$, produces

$$B(h) = -\frac{h^2}{2}\int_z z^2\phi(z)f'(zh^*)$$

where h^* lies between h and zero.

Combining the above arguments,

$$\begin{aligned}|S_n(\boldsymbol{\beta}) - \tilde{S}_n(\boldsymbol{\beta})| &\leq Mn^{1/2}|U_1(h) - U_2(0)| + Mn^{1/2}|B(h)| \\ &= O_p\left(\left[h\log n \ \log\left(\frac{1}{h\log n}\right)\right]^{1/2} + n^{1/2}h^2\right).\end{aligned}$$

Using the Lipschitz and boundedness conditions, along with $nh^4 \to 0$, it follows that

$$S_n(\boldsymbol{\beta}) - \tilde{S}_n(\boldsymbol{\beta}) = o_p(1) \qquad \text{uniformly in } \boldsymbol{\beta}.$$

b) $\hat{\boldsymbol{\beta}}$ *is a consistent estimate of* $\boldsymbol{\beta}_0$

At the first Newton-Raphson iteration,

$$\hat{\boldsymbol{\beta}}^{(1)} = \hat{\boldsymbol{\beta}}_I + \{n^{-1/2}A_n(\hat{\boldsymbol{\beta}}_I)\}^{-1}n^{-1/2}S_n(\hat{\boldsymbol{\beta}}_I),$$

where $\hat{\boldsymbol{\beta}}_I$ is a consistent initial estimate of $\boldsymbol{\beta}_0$.

Denoting the limiting value of $n^{-1/2}S_n(\boldsymbol{\beta})$ by $S(\boldsymbol{\beta})$, the triangle inequality shows

$$|n^{-1/2}S_n(\hat{\boldsymbol{\beta}}_I) - S(\boldsymbol{\beta}_0)| \leq |n^{-1/2}S_n(\hat{\boldsymbol{\beta}}_I) - S(\hat{\boldsymbol{\beta}}_I)| + |S(\hat{\boldsymbol{\beta}}_I) - S(\boldsymbol{\beta}_0)|.$$

By the boundedness conditions, $n^{-1/2}S_n(\boldsymbol{\beta})$ converges uniformly to $S(\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\beta}_0)$ and so for n large, the first term on the right hand side is $\epsilon/2$. The continuity of S and the consistency of $\hat{\boldsymbol{\beta}}_I$ implies that the second term on the right hand side is also $\epsilon/2$ for n large. Since S_n is a U-statistic, $S(\boldsymbol{\beta}_0) = 0$, and so $|n^{-1/2}S_n(\hat{\boldsymbol{\beta}}_I)| < \epsilon$ for n large. Therefore, since $n^{-1/2}A_n(\boldsymbol{\beta}) \neq 0$ for $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\beta}_0)$, $|\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}_I| < \epsilon$ for all $n \geq N$. Repeating these arguments for all iterations until convergence, $\hat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$.

c) The asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$

Taylor expanding $S_n(\hat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}_0$

$$S_n(\hat{\boldsymbol{\beta}}) = S_n(\boldsymbol{\beta}_0) + A_n(\boldsymbol{\beta}^{\#})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

where $\boldsymbol{\beta}^{\#}$ lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Since $n^{-1/2}A_n(\boldsymbol{\beta})$ is bounded and non-zero, it follows from the Lemma and parts a) and b) that $n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}VA^{-1})$, where $V = n^{-1}\text{var}\{S_n(\boldsymbol{\beta}_0)\}$ and A is the limiting value of $n^{-1/2}A_n(\boldsymbol{\beta}_0)$.

References

P. J. Bickel, F. Gotze, and W. R. van Zwet, "The Edgeworth expansion for U-statistics of degree two," *Annals of Statistics* vol. 14 pp. 1463-1484, 1986.

G. E. P. Box and D. R. Cox, " An analysis of transformations," *Journal of Royal Statistical Society, Series B* vol. 26 pp. 211-246, 1964.

J. Buckley and I. R. James, "Linear regression with censored data," *Biometrika* vol. 66 pp. 429-436, 1979.

D. R. Cox, "Regression models and life tables (with Discussion)," *Journal of Royal Statistical Society, Series B* vol. 34 pp. 187-220, 1972.

M. Fygenson and Y. Ritov, "Monotone estimating equations for censored data," *Annals of Statistics* vol. 22 pp. 732-746, 1994.

W. Hardle, P. Hall, and H. Ichimura, "Optimal smoothing in single-index models," *Annals of Statistics* vol. 21 pp. 157-178, 1993.

G. Heller and J. S. Simonoff, "A comparison of estimators for regression with a censored response variable," *Biometrika* vol. 77 pp. 515-520, 1990.

K. R. Hess, D. M. Serachitopol, and B. W. Brown, "Hazard function estimators: a simulation study,". *Statistics in Medicine* vol. 18 pp. 3075-3088, 1999.

M. Joffe, "Administrative and artificial censoring in censored regression models," *Statistics in Medicine* vol. 20 pp. 2287-2304, 2001.

T. L. Lai and Z. Ying, "Large sample theory of a modified Buckley-James estimator for regression analysis with censored data," *Annals of Statistics* vol. 19 pp. 1370-1402, 1991.

R. G. Miller, *Survival Analysis*, Wiley:New York, 1981.

J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal* vol. 7 pp. 308-313, 1965.

D. Oakes, "A note on the Kaplan-Meier estimator," *American Statistician* vol. 47 pp. 39-40, 1993.

Y. Ritov, "Estimation in a linear regression model with censored data," *Annals of Statistics* vol. 18 pp. 303-328, 1990.

J. Robins, "Estimation of the time-dependent accelerated failure time model in the presence of confounding factors," *Biometrika* vol. 79 pp. 321-334, 1992.

J. M. Robins and A. A. Tsiatis, "Correcting for non-compliance in randomized trials using rank preserving structural failure time models," *Communications in Statistics - Theory and Methods* vol. 20 pp. 2609-2631, 1991.

G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, Wiley: New York, 1986.

A. A. Tsiatis, "Estimating regression parameters using linear rank tests for censored data," *Annals of Statistics* vol. 18 pp. 354-372, 1990.

Z. Ying, "A large sample study of rank estimation for censored regression data," *Annals of Statistics* vol. 21 pp. 76-99, 1993.

Z. Ying, S. H. Jung, and L. J. Wei, "Survival analysis with median regression models," *Journal of the American Statistical Association* vol. 90 pp. 178-184, 1995.