Biostatistics Advance Access published May 7, 2014

Biostatistics (2014), *pp.* 1–10
doi:10.1093/biostatistics/kxu021

A Monte Carlo method for variance estimation for estimators based on induced smoothing

ZHEZHEN JIN*
Department of Biostatistics, Columbia University, New York, NY 10032, USA
zj7@columbia.edu

YONGZHAO SHAO
Division of Biostatistics, New York University School of Medicine, New York, NY 10016, USA

ZHILIANG YING
Department of Statistics, Columbia University, New York, NY 10027, USA

Summary

An important issue in statistical inference for semiparametric models is how to provide reliable and consistent variance estimation. Brown and Wang (2005. Standard errors and covariance matrices for smoothed rank estimators. *Biometrika* **92**, 732–746) proposed a variance estimation procedure based on an induced smoothing for non-smooth estimating functions. Herein a Monte Carlo version is developed that does not require any explicit form for the estimating function itself, as long as numerical evaluation can be carried out. A general convergence theory is established, showing that any one-step iteration leads to a consistent variance estimator and continuation of the iterations converges at an exponential rate. The method is demonstrated through the Buckley–James estimator and the weighted log-rank estimators for censored linear regression, and rank estimation for multiple event times data.

Keywords: Accelerated failure time model; Asymptotic fiducial distribution; Buckley–James estimator; Censored data; Contraction mapping; Estimating function; Kaplan–Meier estimator; Monte Carlo integration; Rank estimator.

1. Introduction

Many important estimators involve solving non-smooth and perhaps discontinuous estimating functions. Examples include least absolute deviation estimator (Bloomfield and Steiger, 1983), various rank estimators (Hettmansperger and McKean, 1998), and parameter estimators in censored linear regression (Buckley and James, 1979; Ritov, 1990; Tsiatis, 1990; Lai and Ying, 1991; Ying, 1993). Such estimators are also common in the econometrics literature, where robust procedures are advocated and censoring and sampling bias may occur. Koenker and Bassett (1978) pioneered the quantile regression method and Powell (1984) developed an extended least absolute deviation estimator for censored regression, all involving non-smooth estimating equations.

*To whom correspondence should be addressed.

When the estimating functions are non-smooth, the limiting distributions of the resulting estimators often involve density functions, as exhibited in the above-cited examples. It is therefore desirable to develop methods for variance estimation that bypass density estimation. An interesting development is due to Brown and Wang (2005), where they used a pseudo-Bayesian approach to obtain a naturally induced smoothed version whereby a consistent variance estimator can be obtained through an iterative procedure. They demonstrated usefulness of their method through rank estimators. The work of Brown and Wang (2007) contains an extension of the method to the censored linear regression using the Gehan estimating function. Wang and Zhao (2008), Johnson and Strawderman (2009), and Fu *and others* (2010) further applied the procedure to the analysis of clustered data. Recently, the procedure has been successfully applied to different regression models, Pang *and others* (2012) on the censored quantile regression model, Li *and others* (2012) on the accelerated hazards model, and Lin and Peng (2013) on the linear transformation model.

It appears that a key component in the implementation and theoretic analysis in Brown and Wang (2005, 2007) is that the smoothed version of the estimating function, i.e. integration of the original estimating function with respect to a normal kernel, has a closed analytic form from which the iterative algorithm and convergence analysis can be carried out. This would exclude many well-known estimators of which the estimating functions are non-smooth and non-monotone. In particular, it excludes all weighted log-rank estimators (except for the case of Gehan) and the Buckley–James estimator for censored linear regression. They also noted that their approach is effective only when the underlying estimating function is monotone.

Motivated by Brown and Wang (2005), this paper develops a general way for approximating standard errors. It is based on the use of numerical approximations to integrals and modifies the Brown–Wang method so that the scope and applicability are substantially expanded. This new development is particularly appealing to complicated situations, such as the Buckley–James and weighted log-rank estimators in censored linear regression where the estimating functions take complex forms. The lack of an explicit form for the smoothed estimating function in the general situation also entails that new analytic tools are needed to ensure general convergence of the iterative algorithm. Indeed, we show that the so-called contraction mapping theorem is applicable stochastically and, in consequence, an exponential rate of convergence is established.

The paper is organized as follows. In the next section, we describe our proposed method along with theory. In Section 3, we apply the proposed method to estimate the variance of the parameter estimators in rank estimation and least squares estimation for censored regression, and discuss its extension to the multivariate cases. In Section 4, we present several simulation studies and real examples. We conclude with a discussion in Section 5. Supplementary material available at *Biostatistics* online outlines all theoretical proofs.

2. METHOD AND THEORY

Let θ be a p-dimensional vector of parameters that is related to the observations Y. We shall use $U(\theta) = U(Y, \theta)$ to denote a vector of estimating functions for θ. The resulting estimator $\hat{\theta}$ is obtained by solving $U(\theta) = 0$. Without loss of generality, we assume that U is properly scaled so that $U(\cdot)$ converges to a non-random function $\bar{u}(\cdot)$ and $\bar{u}(\theta_0) = 0$, where θ_0 denotes the true parameter. The first two basic assumptions are as follows.

ASSUMPTION A1 $n^{1/2}U(\theta_0)$ is asymptotically normal with mean 0 and a covariance matrix D, i.e.

$$n^{1/2}U(\theta_0) \Rightarrow N(0, D). \tag{2.1}$$

ASSUMPTION A2 The estimator $\hat{\theta}$ obtained by solving $U(\theta) = 0$ is $\sqrt{n}$-consistent, and $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean 0 and covariance matrix V.

Assumptions A1 and A2 are usually satisfied by many estimating functions. Inference on θ, e.g. construction of a confidence set, requires a consistent estimate of V. As noted in Jin *and others* (2001) and Brown and Wang (2005), a consistent estimator $\hat{D}$ of D is usually easy to obtain because it only involves $U(\theta_0)$. If $U(\theta)$ has a continuous derivative $U'(\theta)$, the typical asymptotic arguments, as discussed in Brown and Wang (2005), would lead to

$$n^{1/2}(\hat{\theta} - \theta_0) = -A^{-1} n^{1/2} U(\theta_0) + o_p(1), \tag{2.2}$$

that is, $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean 0 and covariance matrix $V = A^{-1} D A^{-1}$, where $A = EU'(\theta_0)$. Thus, A can be estimated by $\hat{A} = U'(\hat{\theta})$ and V can be estimated by $\hat{A}^{-1}\hat{D}\hat{A}^{-1}$. However, the estimating function U is often non-smooth, thus one cannot estimate the slope matrix by simply taking partial derivatives. As a consequence, variance estimation for $\hat{\theta}$ can be a challenging issue.

The idea behind the elegant approach of Brown and Wang (2005) is to use the asymptotic fiducial distribution as the basis for an induced smoothing kernel. Specifically, $\hat{\theta} - \theta_0$ is in distribution approximately equal to $n^{-1/2} V^{1/2} Z$, where Z is the standard p-variate normal random vector. It induces the following smoothed version of the estimating function:

$$\tilde{U}(V; \theta) = E_Z U(\theta + n^{-1/2} V^{1/2} Z), \tag{2.3}$$

where E_Z denotes the expectation with respect to Z. Brown and Wang (2005) then suggested to obtain $\hat{\theta}$ and $\hat{V}$ by jointly solving $\tilde{U}(V; \theta) = 0$ and the following equation:

$$V = \tilde{A}^{-1}(V; \hat{\theta}) \hat{D} \tilde{A}^{-1}(V; \hat{\theta}), \tag{2.4}$$

where $\tilde{A}(V; \theta) = \partial \tilde{U}(V; \theta)/\partial \theta^{\mathrm{T}}$ and $\hat{D}$ is a consistent estimator of D. Because both sides of (2.4) involve V, an iterative algorithm for solving V results.

For certain rank estimators and quantiles, Brown and Wang (2005, 2007) were able to obtain manageable analytic forms for $\tilde{U}(V; \theta)$ and $\tilde{A}(V; \theta)$ and showed that their proposed iterative algorithm converges numerically. The approach, however, is not applicable if estimating functions $U(\theta)$ are non-smooth and their smoothed version $\tilde{U}(V; \theta)$ is too complicated to be written out in simple analytic forms. Examples of such kind include all weighted log-rank estimators (except the Gehan estimator) and the Buckley–James estimator for censored linear regression. Thus, there is a need to develop a simple and more generally applicable algorithm to estimate the variance V. It is also desirable to investigate the convergence property of the iterative algorithms under more general conditions.

Next we state another assumption on local asymptotic linearity (LAL) which is generally satisfied even for estimating functions $U(\cdot)$ that are non-smooth and/or non-monotone. In fact, the LAL is a commonly used assumption for proving asymptotic normality. In particular, all examples considered in Brown and Wang (2005, 2007) and in this paper satisfy the following LAL assumption.

ASSUMPTION A3 *U is locally asymptotically linear (LAL) at θ_0, i.e.*

$$\sup_{\theta \in N_0} \|n^{1/2} U(\theta) - n^{1/2} U(\theta_0) - A\sqrt{n}(\theta - \theta_0)\| / (1 + n^{1/2}\|\theta - \theta_0\|) = o_p(1), \tag{2.5}$$

where A is a non-degenerate slope matrix, $\|\cdot\|$ denotes the Euclidean norm, and N_0 is some small neighborhood of θ_0.

Note that Assumptions A1–A3 imply (2.2) with A as defined in (2.5). That is, $n^{1/2}(\hat{\theta}-\theta_0)=-A^{-1}n^{1/2}U(\theta_0)+o_p(1)$. It follows that $n^{1/2}(\hat{\theta}-\theta_0)$ is asymptotically normal with mean 0 and covariance matrix $V=A^{-1}DA^{-1}$.

Using either Stein's Identity (Stein, 1981) or a simple integration by parts argument, the derivative of the smoothed estimating equation $\tilde{U}$ defined in (2.3) satisfies the following equation:

$$A(\Gamma;\theta)=E_Z U(\theta+\Gamma Z)Z^{\mathrm{T}}\Gamma^{-1}. \tag{2.6}$$

REMARK 2.1 The validity of (2.6) does not require the existence of partial derivative of $U(\cdot)$. It is valid as long as the order of the partial derivative $\partial/\partial\theta$ and the expectation E_Z is exchangeable, which can be checked with Fubini's Theorem in measure theory; see supplementary material available at *Biostatistics* online for a proof.

Under the above three general Assumptions A1–A3, a very simple consistent estimate of A is given by $A(\Gamma_0,\hat{\theta})$, where $\Gamma_0=n^{-1/2}I$ and with I being the identity matrix. Note that the two integrals in (2.3) and (2.6) can be numerically approximated arbitrarily well by a simple Monte Carlo method (MCM) or the Gaussian quadrature method (GQM). More specifically, we propose the following two numerical methods to provide simple consistent estimates of the variance V.

An MCM

Step 1: Calculate a consistent estimator $\hat{D}$ of the matrix D.

Step 2: Choose $\Gamma_0=n^{-1/2}I$ and a large number m.

Step 3: For the kth step ($k\geqslant 1$), generate z_j, $j=1,\ldots,m$, from multivariate normal distribution $N(0,I)$. Estimate $\tilde{A}(V;\theta)$ by

$$A_k=A(\Gamma_{k-1})=\frac{1}{m}\sum_{j=1}^{m}U(\hat{\theta}+\Gamma_{k-1}z_j)z_j^{\mathrm{T}}\Gamma_{k-1}^{-1}.$$

Step 4: Calculate $G_k=A_k^{-1}\hat{D}A_k^{-1}$ and define $\Gamma_k=G_k^{1/2}n^{-1/2}$.

Step 5: Repeat Steps 3 and 4 for next k until G_k and Γ_k converge.

The covariance matrix of $\sqrt{n}(\hat{\theta}-\theta_0)$ will be estimated using the G_k at the convergence in the above iterative algorithm.

The convergence of G_k and Γ_k can be assessed by commonly used matrix convergence criteria, such as the difference or relative change. One may choose a very large m to ensure a good approximation for the Gaussian integral.

A GQM

Replace Step 3 in MCM with

Step 3′: Choose grid p-vector points z_j, $j=1,\ldots,m$, based on a pre-specified accuracy criterion, and calculate

$$A_k=A(\Gamma_{k-1})=\sum_{j=1}^{m}U_n(\hat{\theta}+\Gamma_{k-1}z_j)z_j^{\mathrm{T}}w_j\Gamma_{k-1}^{-1},$$

where w_j, $j=1,\ldots,m$, are the Gaussian quadrature weights. One choice of $\{z_j\}$ is based on 1D Gauss–Hermite quadrature calculations.

The following theorem justifies the convergence of the algorithms in MCM and GQM. A proof of the theorem can be found in supplementary material available at *Biostatistics* online.

THEOREM 2.2 Under Assumptions A1–A3, the one-step ($k = 1$) estimates of A and V in the MCM or in the GQM with a large m are consistent as $n \to \infty$. Moreover, the iteration algorithm in either MCM or GQM converges under Assumptions A1–A3.

3. EXAMPLES

We illustrate the methods with three examples: weighted log-rank estimators, the Buckley–James estimators for censored linear regression, and their extensions to multivariate data. It should be noted that the approach of Brown and Wang (2005, 2007) is not applicable to any of the three examples as the corresponding estimating functions are non-monotone and do not have simple form to give an explicit evaluation for the induced smooth versions.

3.1 *Weighted log-rank estimators*

Consider the accelerated failure time (AFT) model for survival times (Kalbfleisch and Prentice, 2002). Let T_i be the failure time and X_i be the p-vector of covariates for the ith individual, $i = 1, \dots, n$. The AFT model relates the logarithm of the failure time, $Y_i = \log T_i$, linearly to the covariates

$$Y_i = X_i^{\mathrm{T}} \beta_0 + \epsilon_i, \tag{3.1}$$

where ϵ_i are independent and identically distributed random errors with unknown distribution function $F(\cdot)$ and β_0 is a p-vector of unknown regression parameters. Instead of the T_i, we observe $\tilde{T}_i = \min\{T_i, C_i\}$ and $\delta_i = 1\{T_i \leqslant C_i\}$ where C_i are censoring times. We assume that observations $(\tilde{T}_i, \delta_i, X_i)$, $i = 1, \dots, n$, are independent and identically distributed.

The general weighted log-rank estimating function with weight function $\psi(\cdot)$ takes the form

$$U(\beta) = \sum_{i=1}^{n} \psi_i(\beta)\delta_i \left(X_i - \frac{\sum_{j=1}^{n} X_j 1\{e_j(\beta) \geqslant e_i(\beta)\}}{\sum_{j=1}^{n} 1\{e_j(\beta) \geqslant e_i(\beta)\}} \right), \tag{3.2}$$

where $e_i(b) = \tilde{Y}_i - X_i^{\mathrm{T}} b$. The corresponding estimate solves $U(\beta) = 0$.

In general, the $U(\beta)$ can be neither smooth nor monotone. Step 1 in the previous section can be easily implemented as the variance of $U(\beta_0)$; D can be estimated through the usual plug-in estimator

$$\hat{D} = \frac{1}{n} \sum_{i=1}^{n} \{\psi_i(\hat{\beta})\}^2 \delta_i \left(\frac{\sum_{j=1}^{n} 1\{e_j(\hat{\beta}) \geqslant e_i(\hat{\beta})\} X_j^{\otimes 2}}{\sum_{j=1}^{n} 1\{e_j(\hat{\beta}) \geqslant e_i(\hat{\beta})\}} - \left[\frac{\sum_{j=1}^{n} 1\{e_j(\hat{\beta}) \geqslant e_i(\hat{\beta})\} X_j}{\sum_{j=1}^{n} 1\{e_j(\hat{\beta}) \geqslant e_i(\hat{\beta})\}} \right]^{\otimes 2} \right),$$

where $A^{\otimes 2} = AA^{\mathrm{T}}$ for a matrix A; cf. Ying *and others* (1992). Steps 2–5 can also be implemented straightforwardly.

3.2 *Buckley–James estimators*

For the censored linear regression model in Section 3.1, Buckley and James (1979) considered an estimating equation based on the least squares principle with

$$U(\beta) = \sum_{i=1}^{n} (X_i - \bar{X})(\hat{Y}_i(\beta) - X_i^{\mathrm{T}} \beta),$$

where $\hat{Y}_i(\beta) = \delta_i \tilde{Y}_i + (1-\delta_i)(\int_{e_i(\beta)}^{\infty} u \, \mathrm{d}\hat{F}_\beta(u)/(1-\hat{F}_\beta(e_i(\beta))) + X_i^{\mathrm{T}}\beta)$, and $\hat{F}_\beta$ is the left-continuous version of the Kaplan–Meier estimate of $F(\cdot)$ based on $(e_i(\beta), \delta_i)$. The estimator $\hat{\beta}$ can be obtained by the method of Jin *and others* (2006a).

Step 1 in Section 2 can be easily implemented as the variance of $U(\beta_0)$; D can be estimated by the method in Ying *and others* (1992).

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

The remaining steps are straightforward.

3.3 *Rank estimation for multiple event times data*

Jin *and others* (2006b) considered the extension of the rank estimation to multiple events data, recurrent events data, and clustered failure time data and developed resampling approaches to estimate the limiting covariance matrices without non-parametric density estimation or evaluation of numerical derivatives. However, their implementation is computationally intensive. The approach developed in this paper offers a rather simple way of estimating the limiting covariance matrices. We illustrate the use of the proposed method with the rank estimation in multiple events data.

Suppose that a subject can potentially experience K types of events. For the ith subject, $i = 1, \ldots, n$ and $k = 1, \ldots, K$, let T_{ki} be the time to the kth event, C_{ki} be the corresponding censoring time, and X_{ki} be the corresponding $p_k \times 1$ vector of covariates. The observed data consist of $(\tilde{T}_{ki}, \Delta_{ki}, X_{ki})$ $(k = 1, \ldots, K;\ i = 1, \ldots, n)$, where $\tilde{T}_{ki} = T_{ki} \wedge C_{ki}$ and $\Delta_{ki} = I(T_{ki} \leqslant C_{ki})$.

Jin *and others* (2006b) considered the marginal distributions of the K types of events with AFT models while leaving the dependence structures unspecified.

$$\log T_{ki} = \beta_k^{\mathrm{T}} X_{ki} + \epsilon_{ki}, \quad i = 1, \ldots, n; \quad k = 1, \ldots, K,$$

where β_k is a $p_k \times 1$ vector of unknown regression parameters, and $(\epsilon_{1i}, \ldots, \epsilon_{Ki})$ $(i = 1, \ldots, n)$ are independent random vectors with a common, but completely unspecified, joint distribution that are independent of the X_{ki}.

Let $e_{ki}(b) = \log \tilde{T}_{ki} - b^{\mathrm{T}} X_{ki}$ and $S_k^{(r)}(b; t) = n^{-1} \sum_{i=1}^{n} I\{e_{ki}(b) \geqslant t\} X_{ki}^r (r = 0, 1)$. The weighted log-rank estimating function for β_k is given by

$$\mathbf{U}_{k,\phi_k}(b) = \sum_{i=1}^{n} \Delta_{ki} \phi_k(b; e_{ki}(b)) \{X_{ki} - \bar{X}_k(b; e_{ki}(b))\},$$

where $\bar{X}_k(b; t) = S_k^{(1)}(b; t)/S_k^{(0)}(b; t)$, and ϕ_k is a weight function. The resulting estimator is denoted by $\hat{\beta}_{k,\phi_k}$. Note that the choices of $\phi_k = 1$, $\phi_k = S_k^{(0)}(b; t)$, and ϕ_k being the Kaplan–Meier estimator based on $\{e_{ki}(b), \Delta_{ki}\}$ $(i = 1, \ldots, n)$ as $\phi_k(b; t)$ correspond to the log-rank, Gehan–Wilcoxon, and Prentice–Wilcoxon statistics, respectively.

Let $\mathbf{B} = (\beta_1^{\mathrm{T}}, \ldots, \beta_K^{\mathrm{T}})^{\mathrm{T}}$ and $\hat{\mathbf{B}} = (\hat{\beta}_{1,\phi_1}^{\mathrm{T}}, \ldots, \hat{\beta}_{K,\phi_K}^{\mathrm{T}})'$. The random vector $n^{1/2}(\hat{\mathbf{B}} - \mathbf{B})$ is asymptotically zero-mean normal with covariance matrix $\{\mathbf{A}_{k,\phi_k}^{-1} \mathbf{V}_{kl,\phi_k\phi_l} \mathbf{A}_{l,\phi_l}^{-1};\ k, l = 1, \ldots, K\}$.

Table 1. *Summary statistics for the simulation studies normal error*

Parameter	Censoring (%)	Gehan estimator				Log-rank				Least squares			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
β_1	0	0.002	0.207	0.209	0.945	0.006	0.222	0.225	0.941	0.003	0.202	0.213	0.957
	25	−0.001	0.230	0.227	0.947	0.001	0.246	0.245	0.940	0.000	0.226	0.235	0.949
	50	−0.000	0.257	0.260	0.949	−0.003	0.274	0.280	0.948	−0.002	0.254	0.267	0.953
β_2	0	−0.005	0.213	0.211	0.943	−0.008	0.223	0.227	0.955	−0.003	0.208	0.215	0.954
	25	−0.000	0.233	0.231	0.938	−0.004	0.242	0.247	0.951	0.000	0.226	0.239	0.959
	50	0.005	0.263	0.267	0.955	−0.001	0.277	0.284	0.948	0.003	0.258	0.274	0.963

Bias, bias of the parameter estimator; SE, standard error of the parameter estimator; SEE, mean of the standard error estimator; CP, coverage probability of the 95% confidence interval.

The $\mathbf{V}_{kl,\phi_k\phi_l}$ can be estimated by the empirical estimator of covariance matrix between $\mathbf{U}_{k,\phi_k}(\hat{\beta}_{k,\phi_k})$ and $\mathbf{U}_{l,\phi_l}(\hat{\beta}_{l,\phi_l})$.

Let $\mathbf{U} = (\mathbf{U}_{1,\phi_1}^{\mathrm{T}}, \dots, \mathbf{U}_{K,\phi_K}^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{A} = \mathrm{diag}(\mathbf{A}_{1,\phi_1}, \dots, \mathbf{A}_{K,\phi_K})$. Denote the empirical estimator of covariance matrix of $\mathbf{U}$ as $\hat{\mathbf{V}}$; then the $\mathbf{A}$ can be estimated as follows.

Step 1: Generate $\mathbf{z}_j$, $j = 1, \dots, m$, from $\sum_{k=1}^{K} p_k$-dimensional multivariate normal distribution $N(0, I)$.

Step 3: Choose $\Gamma_0 = n^{-1/2} I$. Then estimate $\mathbf{A}$ by

$$A_0 = \frac{1}{m} \sum_{j=1}^{m} \mathbf{U}(\hat{\mathbf{B}} + \Gamma_0 z_j) z_j^{\mathrm{T}} \Gamma_0^{-1}.$$

Step 4: Calculate $G_1 = A_0^{-1} \hat{V} A_0^{-1}$ and denote $\Gamma_1 = G_1^{1/2}$.

Step 5: Replace Γ_0 with Γ_1; then iterate between Steps 3 and 4 until Γ_k converges.

The covariance matrix of $\hat{\mathbf{B}}$ will be $\hat{\mathbf{A}}_k = \Gamma_k^{\otimes 2}$.

4. Simulations and application to real data

Simulation studies were conducted to assess the performance of the proposed methods. Here we present simulation results for censored linear regression model (3.1) using the Gehan, the log-rank, and the Buckley–James least squares estimating equations. Following Jin *and others* (2006a), we generate failure times from the model

$$\log T = 2 + X_1 + X_2 + \epsilon,$$

where X_1 is Bernoulli with success probability 0.5, X_2 is normal with mean 0 and standard deviation 0.5, and ϵ has the standard normal, extreme value. The censoring times were generated from the uniform Un[0, c] distribution, where c was chosen to yield a desired level of censoring. We estimated β_1 and β_2 with the log-rank estimation method as in Jin *and others* (2003) and the least squares method as in Jin *and others* (2006a). The 1000 Monte Carlo standard 2D normal random vectors were used and Γ_0 was set to be $n^{-1/2} I$.

The results for a sample size of 100 based on 1000 simulated datasets are summarized in Tables 1 and 2. In all cases, the proposed procedure accurately estimates the variability of the parameter estimator, and the confidence intervals have proper coverage probabilities.

We applied the method to the data on multiple myeloma reported by Krall *and others* (1975), which is the main example in SAS PROC PHREG. (SAS Institute, 1999). Two standardized covariates log(BUN)

Table 2. *Summary statistics for the simulation studies extreme-value error*

Parameter	Censoring (%)	Gehan estimator				Log-rank				Least squares			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
β_1	0	−0.002	0.236	0.238	0.953	−0.002	0.204	0.211	0.952	−0.006	0.262	0.278	0.957
	25	0.000	0.289	0.282	0.941	−0.001	0.243	0.247	0.957	−0.003	0.306	0.307	0.949
	50	0.000	0.368	0.362	0.951	−0.001	0.318	0.321	0.956	−0.004	0.379	0.370	0.945
β_2	0	0.005	0.243	0.241	0.945	−0.001	0.216	0.213	0.938	0.005	0.266	0.281	0.956
	25	0.003	0.286	0.288	0.945	−0.007	0.249	0.250	0.949	0.002	0.300	0.314	0.957
	50	0.008	0.361	0.373	0.945	0.004	0.318	0.324	0.947	0.007	0.366	0.381	0.955

Bias, bias of the parameter estimator; SE, standard error of the parameter estimator; SEE, mean of the standard error estimator; CP, coverage probability of the 95% confidence interval.

and hemoglobin at diagnosis (HGB) were considered for the censored regression model in Section 3.1. The 10 000 Monte Carlo standard 2D normal random vectors were used, Γ_0 was set to be $n^{-1/2}I$, and 0.0001 was used as the convergence criterion between successive estimates. Convergence was reached after three or four iterations. It yielded standard errors (0.142, 0.168) for the Gehan estimate (−0.532, 0.292), (0.173, 0.158) for the log-rank estimate (−0.505, 0.268), and (0.122, 0.146) for the least-squares estimate (−0.520, 0.281). The results are similar to those obtained with the resampling approach in Jin *and others* (2003, 2006a).

We also did reanalysis of the Stanford heart transplantation data in Miller and Halpern (1982) by regressing the base-10 logarithm of the survival time on the patient's age and the T5 mismatch score for the 157 patients with complete records on the T5 mismatch score; the 10 000 Monte Carlo standard 2D normal random vectors were used, Γ_0 was set to be $n^{-1/2}I$ and 0.0001 was used as the convergence criterion between successive estimates. After convergence in four iterations, it yielded standard errors (0.0090, 0.1565) for the Gehan estimate (−0.0211, −0.0265) and (0.0089, 0.1540) for the least-squares estimate (−0.0148, −0.0028). The results are similar to those obtained with the resampling approach in Jin *and others* (2006a).

In our numerical studies, the proposed MCM used significantly less computational time compared with the resampling method for a similar accuracy in results.

5. DISCUSSION

Variance estimation is an important aspect in semiparametric inference. It can be a thorny issue when the corresponding estimating functions are non-smooth. The Brown–Wang approach provides a simple solution through an induced smoothing, and can be easily implemented and justified when the closed form of the induced smoothed estimating function is available.

The present paper expands the scope and applicability of the Brown–Wang approach by recognizing that smoothing can be carried out via Monte Carlo approximations. This is especially crucial when the underlying estimating equations involve the empirical version of the infinite-dimensional parameter in the semiparametric model, as being demonstrated through several examples that are common in semiparametric analysis of failure time data.

The paper focuses on the parametric component of the semiparametric model. It is certainly of interest to extend the approach so that inference for the non-parametric component can be carried out properly. This may require effective handling of estimation for the non-parametric part without creating a large number of estimating equations that increases with the sample size.

Supplementary material

Supplementary Material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

We thank an associate editor and two referees for their careful reading and valuable comments. *Conflict of Interest*: None declared.

Funding

Y.S.'s research is partially supported by the NYU Cancer Center Support Grant 2P30 CA16087.

References

Bloomfield, P. and Steiger, W. L. (1983). *Least Absolute Deviations. Theory, Applications, and Algorithms*. Progress in Probability and Statistics 6. Boston, MA: Birkhäuser Boston.

Brown, B. M. and Wang, Y.-G. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika* **92**, 732–746.

Brown, B. M. and Wang, Y.-G. (2007). Induced smoothing for rank regression with censored survival times. *Statistics in Medicine* **26**, 828–836.

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

Fu, L., Wang, Y. and Bai, Z. (2010). Rank regression for analysis of clustered data: a natural induced smoothing approach. *Computational Statistics and Data Analysis* **54**, 1036–1050.

Hettmansperger, T. P. and McKean, J. W. (1998) *Robust Nonparametric Statistical Methods*. London: Arnold.

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

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

Jin, Z., Lin, D. Y. and Ying, Z. (2006b). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scandinavian Journal of Statistics* **33**, 1–23.

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

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

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. Hoboken: John Wiley.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

Krall, J. M., Uthoff, V. A. and Harley, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics* **31**, 49–57.

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

Li, H., Zhang, J. and Tang, Y. (2012). Induced smoothing for the semiparametric accelerated hazards model. *Computational Statistics and Data Analysis* **56**, 4312–4319.

LIN, H. AND PENG, H. (2013). Smoothed rank correlation of the linear transformation regression model. *Computational Statistics and Data Analysis* **57**, 615–630.

MILLER, R. AND HALPERN, J. (1982). Regression with censored data. *Biometrika* **69**, 521–531.

PANG, L., LU, W. AND WANG, H. J. (2012). Variance estimation in censored quantile regression via induced smoothing. *Computational Statistics and Data Analysis* **56**, 785–796.

POWELL, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* **25**, 303–325.

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics* **18**, 303–328.

SAS INSTITUTE. (1999). *SAS/STAT User's Guide*, Version 8. Cary NC: SAS Institute Inc.

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics* **9**, 1135–1151.

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* **18**, 354–372.

WANG, Y.-G. AND ZHAO, Y. (2008). Weighted rank regression for clustered data analysis. *Biometrics* **64**, 39–45.

YING, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics* **21**, 76–99.

YING, Z., WEI, L. J. AND LIN, J. S. (1992). Prediction of survival probability based on a linear regression model. *Biometrika* **79**, 205–209.

[*Received December 13, 2013; revised March 28, 2014; accepted for publication March 31, 2014*]