Published in final edited form as:
Biostatistics. 2008 April ; 9(2): 355–363. doi:10.1093/biostatistics/kxm034.

Efficient resampling methods for nonsmooth estimating functions

DONGLIN ZENG and **D. Y. LIN**
Department of Biostatistics, CB 7420, University of North Carolina, Chapel Hill, NC 27599-7420, USA lin@bios.unc.edu

Summary

We propose a simple and general resampling strategy to estimate variances for parameter estimators derived from nonsmooth estimating functions. This approach applies to a wide variety of semiparametric and nonparametric problems in biostatistics. It does not require solving estimating equations and is thus much faster than the existing resampling procedures. Its usefulness is illustrated with heteroscedastic quantile regression and censored data rank regression. Numerical results based on simulated and real data are provided.

Keywords

Bootstrap; Censoring; Quantile regression; Rank regression; Robustness; Variance estimation

1. Introduction

The parameters of interest in biostatistics are typically estimated by minimizing a loss function or more generally by solving an estimating equation. In many nonparametric and semiparametric situations, such as Huber's (1964) robust estimation of location (with nonsmooth loss functions), quantile regression, and rank regression, the estimating functions are not differentiable. Then, the asymptotic variances of the parameter estimators generally involve unknown density functions and are thus difficult to estimate directly.

In such situations, it is natural to appeal to resampling techniques. The familiar bootstrap (Efron and Tibshirani, 1993) estimates variances by resampling from the empirical distribution function. This approach needs to be justified on a case-by-case basis and may not be appropriate in complex situations. Parzen *and others* (1994) proposed a resampling technique by equating the observed data estimating function to a random vector which generates the asymptotic distribution of the estimating function. This technique has been applied to numerous biostatistical problems (e.g. Yao *and others*, 1998; Chen and Jewell, 2001; Cai *and others*, 2006). Hu and Kalbfleisch (2000) provided a similar procedure for linear estimating functions with independent terms by bootstrapping the individual terms. For estimators that can be written as minimizers of certain U-statistics, Jin *and others* (2001) developed a resampling approach by incorporating suitable random variables into the minimand. Their approach was adapted by Jin *and others* (2003, 2006) to the rank and least squares regression with censored data.

Correspondence to: D. Y. LIN.

For permissions, please e-mail: journals.permissions@oxfordjournals.org.

Conflict of Interest: None declared.

All the aforementioned resampling procedures require solving the perturbed estimating equations or minimizing the perturbed loss functions a large number of times. This is computationally very demanding, especially for complex nonlinear functions. In addition, the perturbed estimating equations or loss functions tend to be associated with extreme solutions and are thus unstable. As a result, nonsmooth estimating functions are rarely used in practice.

In the present paper, we propose a new resampling strategy to estimate asymptotic variances of parameter estimators obtained from general nonsmooth estimating functions. Our approach only requires generation of random numbers and evaluation of estimating functions. It does not involve solving any perturbed estimating equations or minimizing any perturbed objective functions; therefore, it is far more efficient and more stable than the existing resampling methods. With our approach, variance estimation for complex nonsmooth estimating functions can be accomplished in a matter of seconds or minutes rather than hours or days. We describe the proposed approach in Section 2. We present simulation results and medical examples in Sections 3 and 4, respectively. We provide some concluding remarks in Section 5.

2. Methods

Let θ_0 denote a d-vector of parameters. We estimate θ_0 by solving the estimating equation $U_n(\theta) = 0$, where U_n is a function based on n independent observations such that $n^{-1}U_n(\theta_0) \rightarrow_p 0$. Suppose that the solution $\hat{\theta}$ exists and is consistent. Suppose also that, uniformly in a neighborhood of θ_0,

$$n^{-1/2}U_n(\theta)=n^{-1/2}\sum_{i=1}^{n}S_i+An^{1/2}(\theta-\theta_0)+\mathrm{o}_p(1+n^{1/2}\parallel\theta-\theta_0\parallel), \tag{2.1}$$

where S_i ($i = 1$, mldr;, n) are independent zero-mean random vectors, and A is a nonsingular matrix, which is the asymptotic slope of $n^{-1}U_n(\theta_0)$. This asymptotic expansion holds for a wide variety of estimating functions and can typically be verified through empirical process arguments (van der Vaart and Wellner, 1996, Section 3.3). The S_i are the influence functions for $U_n(\theta_0)$. The dependence of S_i and A on θ_0 is suppressed. Since $U_n(\hat{\theta}) = 0$ and $\hat{\theta}$ is consistent, (2.1) implies that $\hat{\theta}$ is $n^{1/2}$-consistent and $n^{1/2}(\hat{\theta}-\theta_0)$ is asymptotically zero-mean normal with covariance matrix $A^{-1}V(A^{-1})^{\mathrm{T}}$, where $V=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}S_iS_i^{\mathrm{T}}$. For parametric likelihood, $U_n(\theta_0)=\sum_{i=1}^{n}S_i$ and $V=-A$, where S_i is the score for the ith observation and A is the negative information matrix.

We give 2 examples.

Example 1 (Heteroscedastic quantile regression)

For $i = 1$, mldr;, n, let Y_i and X_i denote the response variable and a set of covariates for the ith subject. Assume that the 100τth percentile of Y_i is $\alpha_0+\beta_0^{\mathrm{T}}X_i$. We may estimate $\theta_0 \equiv (\alpha_0,\beta_0^{\mathrm{T}})^{\mathrm{T}}$ by solving the equation

$$\sum_{i=1}^{n}\{I(Y_i-\alpha-\beta^{\mathrm{T}}X_i\le 0)-\tau\}(1,X_i^{\mathrm{T}})^{\mathrm{T}}=0,$$

where $I(\cdot)$ is the indicator function. The solution $\hat{\theta}$ can be obtained by minimizing the loss function

$$\sum_{i=1}^{n}\rho_\tau(Y_i-\alpha-\beta^{\mathrm{T}}X_i),$$

where $\rho_\tau(\upsilon)$ is $\tau\upsilon$ if $\upsilon > 0$ and $(\tau-1)\upsilon$ if $\upsilon \le 0$. This minimization can be performed by linear programing (Koenker and D'Orey, 1987). Under the assumption that $(Y_i-\alpha_0-\beta_0^{\mathrm{T}}X_i)$ has a unique 100τth percentile at 0 and has a continuous density function f_i such that $f_i(0)$ is strictly positive, the estimator $\hat{\theta}$ is consistent and the asymptotic expansion (2.1) holds with $S_i=\{I(Y_i-\alpha_0-\beta_0^{\mathrm{T}}X_i\le 0)-\tau\}(1,X_i^{\mathrm{T}})^{\mathrm{T}}$ (Jin *and others*, 2001). The slope matrix A involves the density functions $f_{i.}$ Buchinsky (1995) compared various bootstrap procedures for estimating the asymptotic covariance matrix of $\hat{\theta}$.

Example 2 (Rank regression with censored data)

Assume that

$$Y_i=\beta_0^{\mathrm{T}}X_i+\varepsilon_i, \tag{2.2}$$

where ε_i (i = 1, mldr;, n) are independent and identically distributed random variables that are independent of X_i (i = 1, mldr;, n). Suppose that Y_i is subject to censoring by $C_{i.}$ In survival analysis, Y_i and C_i are usually expressed on the log-scale and (2.2) is referred to as the accelerated life or accelerated failure time model (Cox and Oakes, 1984, pp. 64–65; Kalbfleisch and Prentice, 2002, pp. 218–219). The data consist of $(\tilde{Y}_i, \Delta_i, X_i)$ (i = 1, mldr;, n), where $\tilde{Y}_i = \min(Y_i, C_i)$ and $\Delta_i = I(Y_i \le C_i)$. It is assumed that C_i is independent of Y_i conditional on $X_{i.}$ One may estimate β_0 by the log-rank estimating equation

$$\sum_{i=1}^{n}\Delta_i\left\{X_i-\frac{\sum_{j=1}^{n}I(\tilde{Y}_j-\beta^{\mathrm{T}}X_j\ge\tilde{Y}_i-\beta^{\mathrm{T}}X_i)X_j}{\sum_{j=1}^{n}I(\tilde{Y}_j-\beta^{\mathrm{T}}X_j\ge\tilde{Y}_i-\beta^{\mathrm{T}}X_i)}\right\}=0. \tag{2.3}$$

It is not a trivial matter to solve this discrete equation, especially when d is large. One may use bisection search or optimization algorithms, such as simulated annealing (Lin and Geyer, 1992). Recently, Jin *and others* (2003) showed that linear programing can be used to obtain an approximation to the log-rank estimate. Under mild conditions (Tsiatis, 1990; Ying, 1993), expansion (2.1) holds with

$$S_i=\Delta_i\left\{X_i-\frac{\Gamma_1(\tilde{Y}_i-\beta_0^{\mathrm{T}}X_i)}{\Gamma_0(\tilde{Y}_i-\beta_0^{\mathrm{T}}X_i)}\right\}-\int_{-\infty}^{\tilde{Y}_i-\beta_0^{\mathrm{T}}X_i}\left\{X_i-\frac{\Gamma_1(t)}{\Gamma_0(t)}\right\}\mathrm{d}\Lambda_0(t),$$

where

$$\Gamma_0(t)=\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} I(\tilde{Y}_i-\beta_0^{\mathrm{T}}X_i\ge t),\ \Gamma_1(t)=\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} I(\tilde{Y}_i-\beta_0^{\mathrm{T}}X_i\ge t)X_i,$$

and Λ_0 is the cumulative distribution function of $\varepsilon_{i.}$ In this case, direct estimation of A would require estimation of the hazard function or density function of ε_i.

It is natural to estimate V directly by $\widehat{V}\equiv n^{-1}\sum_{i=1}^{n}\widehat{S}_i\widehat{S}_i^{\mathrm{T}}$,where $\hat{S}_i$ is obtained from S_i by replacing the unknown quantities by their sample estimators. In Example 1, only θ_0 is unknown; in Example 2, the unknown quantities include β_0, $\Gamma_0(\cdot)$, $\Gamma_1(\cdot)$, and $\Lambda_0(\cdot)$. The consistency of $\hat{V}$ can typically be established by empirical process arguments.

When the $\hat{S}_i$ have complicated expressions, it is more convenient and perhaps more accurate to bootstrap from the data. Let $U_n^*(\theta)$ denote the estimating function based on the bootstrap sample. It follows from (2.1) that

$$n^{-1/2}U_n^*(\theta)=n^{-1/2}\sum_{i=1}^{n}M_iS_i+An^{1/2}(\theta-\theta_0)+\mathrm{o}_p(1+n^{1/2}\parallel\theta-\theta_0\parallel),$$

where M_i is the number of times the ith observation appears in the bootstrap sample. Since $U_n(\hat{\theta})=0$ by definition, we obtain

$$n^{-1/2}U_n^*(\widehat{\theta})=n^{-1/2}U_n^*(\widehat{\theta})-n^{-1/2}U_n(\widehat{\theta})=n^{-1/2}\sum_{i=1}^{n}(M_i-1)S_i+\mathrm{o}_p(1+n^{1/2}\parallel\widehat{\theta}-\theta_0\parallel).$$

By Lemma 3.6.15 of van der Vaart and Wellner (1996), the conditional distribution of $n^{-1/2}U_n^*(\widehat{\theta})$ given the data is asymptotically zero-mean normal with covariance matrix V provided that the remainder term in the above display is $\mathrm{o}_p(1)$ uniformly in the bootstrap samples. It is straightforward to verify the required condition for Examples 1 and 2. The bootstrap estimator of V is also denoted by $\hat{V}$.

To avoid nonparametric density estimation, we propose efficient resampling procedures to estimate A and consequently the asymptotic covariance matrix of $n^{1/2}(\hat{\theta}-\theta_0)$. Let $\tilde{\theta}=\hat{\theta}+n^{-1/2}Z$, where Z is a zero-mean random vector independent of the data. It follows from (2.1) that

$$n^{-1/2}U_n(\tilde{\theta})-n^{-1/2}U_n(\widehat{\theta})=An^{1/2}(\tilde{\theta}-\widehat{\theta})+\mathrm{o}_p(1).$$

Since $U_n(\hat{\theta})=0$ and $\tilde{\theta}-\hat{\theta}=n^{-1/2}Z$, we have

$$n^{-1/2}U_n(\tilde{\theta})=AZ+\mathrm{o}_p(1). \tag{2.4}$$

Thus, we propose the following resampling procedure based on the least squares.

LS method

<u>Step 1:</u> Generate B realizations of Z, denoted by Z_1, mldr;, Z_B.

<u>Step 2:</u> Calculate $n^{-1/2}\ U_n(\hat{\theta} + n^{-1/2}Z_b)$ (b = 1, mldr;, B).

<u>Step 3:</u> For j = 1, mldr;, d, calculate the least squares estimate of $n^{-1/2}\ U_{jn}(\hat{\theta} + n^{-1/2}Z_b)$ (b = 1, mldr;, B) on Z_b (b = 1, mldr;, B), where U_{jn} denotes the jth component of U_n. Let $\hat{A}$ be the matrix whose jth row is the jth least squares estimate.

<u>Step 4:</u> Estimate the covariance matrix of $n^{1/2}(\hat{\theta} - \theta_0)$ by $\hat{A}^{-1}\hat{V}\ (\hat{A}^{-1})^{\mathrm{T}}$.

In many situations, A is symmetric, in which case a simpler resampling procedure can be obtained. If the covariance matrix of Z is V^{-1}, then (2.4) implies that Cov $(n^{-1/2}\ U_n(\tilde{\theta})|\text{data}) = AV^{-1}\ A^{\mathrm{T}} + o_p(1)$. The inverse of this covariance matrix is equal to $A^{-1}V\ (A^{-1})^{\mathrm{T}}$ when A is symmetric. Thus, we propose the following resampling procedure based on the sample variance of $n^{-1/2}U_n(\tilde{\theta})$.

SV method

<u>Step 1:</u> Generate $\tilde{\theta}_b \equiv \hat{\theta}+ n^{-1/2}Z_b$ (b = 1, mldr;, B), where Z_b is a zero-mean random vector with covariance matrix $\hat{V}^{-1}$.

<u>Step 2:</u> Calculate the sample covariance matrix of $n^{-1/2}U_n(\tilde{\theta}_b)$ (b = 1, mldr;, B) and denote it by $\hat{\Sigma}$.

<u>Step 3:</u> Estimate the covariance matrix of $n^{1/2}(\hat{\theta} - \theta_0)$ by $\hat{\Sigma}^{-1}$.

Unlike the existing resampling methods, the least squares (LS) and sample variance (SV) methods do not require solving estimating equations. This is an important advantage since it is computationally intensive to solve complex nonsmooth estimating equations. Although we have suggested the possible use of bootstrap to estimate V, that procedure is different from bootstrap estimation of the variance of $\hat{\theta}$ and does not involve solving equations.

3. Simulation studies

We conducted extensive simulation studies to assess the performance of the proposed resampling methods. For both the LS and the SV methods, we estimated V either by direct evaluation or by bootstrap. We set Z to $\hat{V}^{-1/2}Z^*$, where Z^* is either a d-variate standard normal random vector or a d-vector of independent centerd Bernoulli random variables with equal probabilities at −1 and 1. Thus, 8 different variants of the methods were considered.

The first set of studies mimics the simulation studies on median regression reported in Section 3 of Parzen *and others* (1994). We generated data from the model $Y_i = X_{1i} + X_{2i} + \varepsilon_{i,}$ where X_{1i} and X_{2i} are independent standard normal and Bernoulli with 0.5 success probability, respectively, and ε_i is normal with mean 0 and variance $|X_{1i}|$. We obtained the parameter estimates through linear programing.

The second set of studies is similar to those of Jin *and others* (2003). We generated survival times from model (2.2) in which X_1 and X_2 are independent Uniform(0, 1) variable and Bernoulli variable with 0.5 success probability, $\beta_0 = (1, -1)^{\mathrm{T}}$, and the error distribution is either extreme-value or zero-mean normal with standard deviation 0.5. We generated censoring times from a uniform distribution to yield a censoring rate of 25%. We obtained the log-rank estimates through bisection search.

The results from the above 2 sets of studies are summarized in Tables 1 and 2. The results of Table 1 pertain to the continuous covariate. Each entry in the tables is based on 10 000 simulated data sets and B = 10 000. Clearly, all 8 variants of the resampling methods work well in that the variance estimators accurately reflect the true variations and the associated confidence intervals have proper coverage probabilities. There are virtually no differences between the LS and SV methods or between the direct and bootstrap estimation of V. For the rank regression under the normal error distribution, the Bernoulli sampling appears to be slightly better than the normal sampling. For median regression, the new resampling method is approximately 100 times faster than bootstrap (with 10 000 resamples); for rank regression, it is approximately 1000 times faster.

4. Applications

4.1 Multiple myeloma study

We applied the proposed resampling methods to a multiple myeloma study (Krall *and others*, 1975). Out of the 65 patients who were treated with alkylating agents, 48 died during the study. Following Jin *and others* (2003), we fitted model (142.2) with hemoglobin and the logarithm of blood urea nitrogen as the covariates by using both the log-rank and the Gehan estimators. The Gehan estimator is obtained by incorporating the weight function $n^{-1}\sum_{j=1}^{n} I(\tilde{Y}_j - \beta^{\mathrm{T}} X_j \geq \tilde{Y}_i - \beta^{\mathrm{T}} X_i)$ into (2.3). We considered the 8 variants of the resampling methods evaluated in the simulation studies. The differences are negligible between the LS and the SV methods and between the direct and the bootstrap methods of estimating V.

The results based on the SV method and direct evaluation of V are shown in Table 3. These results are comparable to those of Jin *and others* (2003) but were obtained with much less time.

4.2 Atherosclerosis Risk in Communities Study

We also applied our methods to the Atherosclerosis Risk in Communities Study (The ARIC Investigators, 1989), which is an epidemiologic cohort study of 15 792 subjects aged 45–64 years to investigate the etiology of atherosclerosis and other diseases. We considered all incident coronary heart disease (CHD) cases occurring between 1987 and 2001. We focused on the Caucasian sample, which consists of 11 526 subjects with 774 cases. We used model (2.2) to study the effects of 5 covariates, including smoking status (ever smoke = 1, never smoke = 0), 2 dummy variables contrasting Minnesota and Washington states to North Carolina, gender (male = 1, female = 0), and standardized age at the baseline, on the time to the occurrence of CHD. For large data sets such as this one, the methods of Jin *and others* (2003, 2006) are not computationally feasible. We used the Nelder–Mead algorithm as implemented in MATLAB to calculate the log-rank and Buckley–James estimates. The results based on the LS and SV methods with direct evaluation of V and 10 000 normal random samples are displayed in Table 4. For comparison, we also report the results of the method of Parzen *and others* (1994) with B = 10 000. The standard error estimates are very similar between the LS and the SV methods, whereas those of the method of Parzen *and others* tend to be slightly larger. The larger standard error estimates by the method of Parzen *and others* are likely due to the unstabilities of the perturbed estimating equations. Indeed, the method of Parzen *and others* produced 7 extreme estimates in the Buckley–James estimation of the gender effect, which were excluded in the standard error calculations. For the new resampling approach, it took approximately 1 and 3 min on an IBM BladeCenter HS20 machine to estimate the standard errors for the log-rank and Buckley–James estimators, respectively, whereas the method of Parzen *and others* consumed 10 and 24 h, respectively.

5. Discussion

The existing resampling methods require solving estimating equations or minimizing loss functions repeatedly, whereas the proposed methods only involve the evaluation of estimating functions. In complex situations, such as rank regression and least squares regression with censored data, the amount of time required to evaluate an estimating function is negligible as compared to solving the corresponding estimating equation. Then, the proposed methods are orders of magnitude faster than the existing resampling methods. Despite the continuing improvement in computer power, this degree of saving is very important, especially for large data sets and for simulation studies. Adopting the proposed resampling procedures will not only enhance the utilities of many existing nonparametric and semiparametric estimators but also facilitate the development and evaluation of new methods for complex biostatistical problems.

The approach of Hu and Kalbfleisch (2000) does not require solving estimating equations repeatedly in order to construct confidence intervals but requires to do so for estimating the variances of parameter estimators. It is restricted to linear estimating functions with independent terms and thus would be applicable to quantile regression, but not to rank regression or Buckley–James estimation.

Our method can be viewed as a version of Monte Carlo numerical differentiation. In contrast to the usual numerical differentiation that uses fixed step sizes, the new method generates random step sizes Z, exploring a broad range of step sizes and producing stable estimates. Numerical results indicate that our method is not sensitive to the choice of the distribution of Z.

The proposed methods have very broad applications and are particularly applicable to the situations in which the method of Parzen *and others* has been used. We have focused our attention on nonsmooth estimating functions. In some situations, the estimating functions are differentiable, but the derivatives are difficult to calculate. Then, the proposed resampling methods would also be appealing.

The results of Section 2 continue to hold if (2.1) is replaced by the more general expansion

$$n^{-1/2}U_n(\theta)=G+An^{1/2}(\theta-\theta_0)+o_p(1+n^{1/2}\parallel\theta-\theta_0\parallel),$$

where G is a zero-mean random vector whose covariance matrix can be consistently estimated. Thus, the proposed resampling methods can be applied to multivariate responses, biased sampling, and time series data among others. Indeed, the $n^{1/2}$ convergence rate is not essential. Furthermore, our approach can potentially be extended to semiparametric situations in which infinite-dimensional parameters are part of θ.

Acknowledgements

The authors thank the reviewers for helpful comments.

Funding

National Institutes of Health.

References

Buchinsky M. Estimating the asymptotic covariance-matrix for quantile regression-models—a Monte Carlo study. Journal of Econometrics 1995;68:303–338.

Cai TX, Pepe MS, Zheng YY, Lumley T, Jenny NS. The sensitivity and specificity of markers for event times. Biostatistics 2006;7:182–197. [PubMed: 16079162]

Chen YQ, Jewell NP. On a general class of semiparametric hazards regression models. Biometrika 2001;88:687–702.

Cox, DR.; Oakes, D. Analysis of Survival Data. London: Chapman and Hall; 1984.

Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. New York: Chapman and Hall; 1993.

Hu F, Kalbfleisch JD. The estimating function bootstrap (with discussion). Canadian Journal of Statistics 2000;28:449–499.

Huber PJ. Robust estimation of a location parameter. The Annals of Mathematical Statistics 1964;35:73–101.

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

Jin Z, Lin DY, Ying Z. On the least squares regression with censored data. Biometrika 2006;93:147–162.

Jin Z, Ying Z, Wei LJ. A simple resampling method by perturbing the minimand. Biometrika 2001;88:381–390.

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. Vol. 2. Hoboken, NJ: Wiley; 2002.

Koenker R, D'Orey V. Computing regression quantiles. Applied Statistics 1987;36:383–393.

Krall JM, Uthoff VA, Harley JB. A step-up procedure for selecting variables associated with survival. Biometrics 1975;31:49–57. [PubMed: 1164538]

Lin DY, Geyer CJ. Computational methods for semiparametric linear regression with censored data. Journal of Computational and Graphical Statistics 1992;1:77–90.

Parzen MI, Wei LJ, Ying Z. A resampling method based on pivotal estimating functions. Biometrika 1994;81:341–350.

The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. American Journal of Epidemiology 1989;129:687–702. [PubMed: 2646917]

Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. The Annals of Statistics 1990;18:354–372.

van der Vaart, AW.; Wellner, JA. Weak Convergence and Empirical Processes. New York: Springer; 1996.

Yao Q, Wei LJ, Hogan JW. Analysis of incomplete repeated measurements with dependent censoring times. Biometrika 1998;85:139–149.

Ying Z. A large sample study of rank estimation for censored regression data. The Annals of Statistics 1993;21:76–99.

Table 1

Simulation results for heteroscedastic median regression

n	Bias	SE		LS method				SV method			
				Normal Z		Bernoulli Z		Normal Z		Bernoulli Z	
				SEE	CP	SEE	CP	SEE	CP	SEE	CP
50	−0.000	0.209	$\hat{V}_1$	0.228	0.957	0.225	0.948	0.215	0.948	0.216	0.943
			$\hat{V}_2$	0.228	0.957	0.224	0.947	0.215	0.948	0.216	0.942
100	0.001	0.147	$\hat{V}_1$	0.152	0.946	0.151	0.937	0.147	0.937	0.147	0.932
			$\hat{V}_2$	0.152	0.945	0.151	0.937	0.146	0.937	0.147	0.932
200	−0.000	0.102	$\hat{V}_1$	0.105	0.947	0.102	0.943	0.102	0.941	0.102	0.939
			$\hat{V}_2$	0.105	0.947	0.104	0.942	0.102	0.941	0.102	0.939

Note: Bias and SE are the bias and standard error of the parameter estimator, respectively; SEE and CP denote the mean of the standard error estimator and the coverage probability of the 95% confidence interval, respectively; $\hat{V}_1$ and $\hat{V}_2$ denote the direct estimation of V and the bootstrap estimation of V, respectively.

Table 2

Simulation results for rank regression with censored data

n		**Bias**	**SE**		**LS method**				**SV method**			
					Normal *Z*		**Bernoulli *Z***		**Normal *Z***		**Bernoulli *Z***	
					SEE	**CP**	**SEE**	**CP**	**SEE**	**CP**	**SEE**	**CP**
Extreme-value error												
100	β_1	0.010	0.428	$\hat{V}_1$	0.428	0.949	0.419	0.941	0.426	0.948	0.419	0.941
				$\hat{V}_2$	0.423	0.947	0.415	0.938	0.421	0.945	0.415	0.938
	β_2	−0.003	0.248	$\hat{V}_1$	0.250	0.948	0.245	0.941	0.248	0.947	0.245	0.941
				$\hat{V}_2$	0.247	0.945	0.243	0.939	0.246	0.943	0.242	0.939
200	β_1	0.000	0.295	$\hat{V}_1$	0.295	0.948	0.292	0.944	0.295	0.947	0.295	0.944
				$\hat{V}_2$	0.293	0.946	0.290	0.943	0.293	0.946	0.290	0.943
	β_2	−0.002	0.170	$\hat{V}_1$	0.173	0.953	0.171	0.950	0.170	0.952	0.171	0.950
				$\hat{V}_2$	0.172	0.951	0.170	0.949	0.171	0.951	0.170	0.949
Normal error												
100	β_1	0.005	0.217	$\hat{V}_1$	0.237	0.963	0.225	0.951	0.235	0.962	0.225	0.951
				$\hat{V}_2$	0.235	0.962	0.222	0.949	0.233	0.961	0.222	0.949
	β_2	−0.001	0.126	$\hat{V}_1$	0.138	0.966	0.130	0.956	0.137	0.965	0.130	0.956
				$\hat{V}_2$	0.136	0.964	0.129	0.954	0.135	0.964	0.129	0.953
200	β_1	0.004	0.153	$\hat{V}_1$	0.160	0.956	0.155	0.950	0.159	0.956	0.155	0.949
				$\hat{V}_2$	0.158	0.955	0.154	0.948	0.158	0.955	0.154	0.948
	β_2	−0.001	0.086	$\hat{V}_1$	0.090	0.961	0.090	0.957	0.092	0.960	0.090	0.957
				$\hat{V}_2$	0.092	0.960	0.089	0.956	0.091	0.960	0.089	0.956

Note: see the note to Table 1.

Table 3

Rank regression analysis of the myeloma data

Covariate	**Estimate**	**Normal *Z***		**Bernoulli *Z***	
		Standard error	**95% interval**	**Standard error**	**95% interval**
Hemoglobin					
Log-rank	0.268	0.164	(−0.055, 0.587)	0.158	(−0.044, 0.576)
Gehan	0.292	0.183	(−0.067, 0.651)	0.176	(−0.054, 0.638)
Blood urea nitrogen					
Log-rank	−0.505	0.162	(−0.827, −0.191)	0.161	(−0.825, −0.193)
Gehan	−0.532	0.154	(−0.834, −0.230)	0.149	(−0.823, −0.241)

Table 4

Accelerated failure time regression for the Atherosclerosis Risk in Communities data

Covariate	Estimate	Standard error estimate		
		LS	SV	Parzen
Smoking status				
Log-rank	−0.411	0.060	0.060	0.060
Buckely–James	−0.363	0.087	0.090	0.092
Minnesota				
Log-rank	0.121	0.064	0.064	0.065
Buckely–James	0.093	0.065	0.065	0.068
Washington				
Log-rank	−0.165	0.061	0.061	0.062
Buckely–James	−0.147	0.067	0.067	0.070
Age				
Log-rank	−0.292	0.028	0.028	0.028
Buckely–James	−0.264	0.055	0.054	0.058
Gender				
Log-rank	−0.893	0.065	0.065	0.067
Buckely–James	−0.842	0.176	0.172	0.195