Science in China Series A: Mathematics
Oct., 2008, Vol. 51, No. 10, 1843–1854
www.scichina.com math.scichina.com
www.springerlink.com

Random weighting method for Cox's proportional hazards model

CUI WenQuan[1†], LI Kai[1], YANG YaNing[1] & WU YueHua[2]

[1] Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026, China
[2] Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada
(email: wqcui@ustc.edu.cn, lkai@mail.ustc.edu.cn, ynyang@ustc.edu.cn, wuyh@mathstat.yorku.ca)

Abstract Variance of parameter estimate in Cox's proportional hazards model is based on asymptotic variance. When sample size is small, variance can be estimated by bootstrap method. However, if censoring rate in a survival data set is high, bootstrap method may fail to work properly. This is because bootstrap samples may be even more heavily censored due to repeated sampling of the censored observations. This paper proposes a random weighting method for variance estimation and confidence interval estimation for proportional hazards model. This method, unlike the bootstrap method, does not lead to more severe censoring than the original sample does. Its large sample properties are studied and the consistency and asymptotic normality are proved under mild conditions. Simulation studies show that the random weighting method is not as sensitive to heavy censoring as bootstrap method is and can produce good variance estimates or confidence intervals.

Keywords: **bootstrap, Cox model, censoring rate, random weighting, consistency, asymptotic normality**

MSC(2000): **62N01, 62N02, 62N03, 62F40, 62G09**

1 Introduction

The Cox regression model and the partial likelihood approach[1, 2] have been studied extensively in literature. It is well known that the maximum partial likelihood estimators of the regression parameters are consistent, asymptotically normal distributed and semiparametrically efficient[3, 4]. Statistical inferences are mainly based on the large sample properties. However, when sample size is small or survival data are heavily censored, the asymptotic results may fail to work properly. Alternative approaches such as bootstrap methods as well as its various modifications have been proposed[5−7]. However, if the event times are heavily censored, bootstrap methods may also fail to work properly since censored times could be repeatedly resampled. In this paper we introduce a random weighting method, which can be used to construct an approximation of the distribution of the maximum partial likelihood estimators of the regression parameters in the proportional hazards model.

Let T_i, $1 \leqslant i \leqslant n$, be the possibly right censored survival times for the i-th subject, with the corresponding indicators δ_i of uncensoring and the possibly time-dependent p-dimensional

Received November 28, 2006; accepted July 30, 2007
DOI: 10.1007/s11425-007-0164-7
† Corresponding author
This work was supported by the National Natural Science Foundation of China (Grant Nos. 10471136, 10671189), PhD Program Foundation of Ministry of Education of China and Foundations from the Chinese Academy of Sciences

covariate vector $Z_i = (Z_{i1}, \dots, Z_{ip})^{\mathrm{T}}$, where A^{T} denotes the transpose of A. The Cox's proportional hazards model[1] assumes that the hazard function for T_i, given Z_i, has the form $\lambda_i(t|Z_i) = \lambda_0(t)\exp(\beta^{\mathrm{T}} Z_i(t))$, where β is a p-dimensional unknown regression parameter vector, and $\lambda_0(t)$ is a completely unspecified baseline hazard function. The corresponding partial likelihood[2] is given by

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(\beta^{\mathrm{T}} Z_i(T_i))}{\sum_{l \in R(T_i)} \exp(\beta^{\mathrm{T}} Z_l(T_l))} \right)^{\delta_i},$$

where $R(t) = \{i : T_i \geqslant t\}$ is the risk set at $t > 0$. Let $\tau = \sup\{t : \Pr(T > t) > 0\}$, the terminal time of the study, and $0 < \tau < \infty$. The maximum partial likelihood estimate $\hat{\beta}_n$ is a solution to the following equation:

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \int_0^{\tau} (Z_i(t) - \bar{Z}(\beta, t)) dN_i(t) = 0, \tag{1.1}$$

where $N_i(t) = I\{T_i \leqslant t, \delta_i = 1\},\ t \in [0, \tau]$,

$$\bar{Z}(\beta, t) = \sum_{i=1}^{n} Z_i(t) Y_i(t) \exp(\beta^{\mathrm{T}} Z_i(t)) \Big/ \sum_{i=1}^{n} Y_i(t) \exp(\beta^{\mathrm{T}} Z_i(t)),$$

$Y_i(t) = I\{T_i \geqslant t\}$ is the at-risk process at $t > 0$ and $I(\cdot)$ is the indicator function. It was shown in literature that $\hat{\beta}_n$ is consistent and asymptotically normally distributed. Variance of $\hat{\beta}_n$ can be obtained by using the asymptotic variance. Alternative methods for estimating the variance of $\hat{\beta}_n$ can be found in literature, including the bootstrap method by Efron[8] and the conditional bootstrap method by Hjort[9].

It is noted that when a bootstrap method is used, the original observations are sampled repeatedly. When censoring is heavy, censored observations may become overwhelmingly dominant in the resampled data and the partial likelihood approach may fail to give a converged solution. If survival data have a high censoring rate, the resampled data may possibly produce the regression coefficient estimation with large variances and therefore result in incorrect estimation of the empirical distribution of the regression parameter estimators. This is because some of the bootstrap samples may be more heavily censored than the original sample does, and consequently, the corresponding bootstrap estimates of the interested parameters display large variabilities. To avoid resampling censored survival times repeatedly, we propose using the random weighting method[10, 11] as an alternative to bootstrap methods. Formally, let $w_i, i = 1, 2, \dots, n$ be independently and identically distributed (i.i.d.) non-negative random variables with $E(w_i) = 1$ and $\mathrm{var}(w_i) = 1$, the random weighting estimation equation is a weighted version of (1.1)

$$U^*(\beta) \triangleq \sum_{i=1}^{n} w_i \int_0^{\tau} (Z_i(t) - \bar{Z}^*(\beta, t)) dN_i(t) = 0, \tag{1.2}$$

with

$$\bar{Z}^*(\beta, t) = \sum_{i=1}^{n} w_i Z_i(t) Y_i(t) \exp(\beta^{\mathrm{T}} Z_i(t)) \Big/ \sum_{i=1}^{n} w_i Y_i(t) \exp(\beta^{\mathrm{T}} Z_i(t)).$$

The random weighting estimate $\hat{\beta}_n^*$ is defined to be a solution to the above equation.

The random weighting method works through repeatedly re-weighting the estimation equation randomly to generate the copies of the original interested parameters or quantities. In this spirit, the random weighting method is similar to bootstrap methods. In fact, the random weighting method[10] was also referred to as the Bayesian bootstrap method[12−14]. This method has been used in many applications as an alternative to bootstrap method. For example, Rao and Zhao[11] used this method in the parameter estimation of a linear regression model. Large sample studies show that the random weighting method has similarly asymptotic properties as bootstrap methods (at least at the first order) (see [10, 11]). An advantage of the random weighting method is that no observation is repeatedly used within each replicate of the random weighting, though each observation may be weighted unequally. This property is especially useful in the analysis of survival data, censored times are regarded as being less informative and too many inclusions could lead to difficulties in obtaining good parameter estimates.

In this paper, we study the applications of the random weighting method in survival analysis, focusing on the large sample theory as well as the applications to survival data with small sample sizes and high censoring rates. We conduct the simulation studies to compare the random weighting method with bootstrap method. It is shown that the random weighting is less sensitive to heavy censoring than bootstrap method does. The rest of this paper is organized as follows. In Section 2, we present the random weighting methods and main results. Simulation results and some discussions are provided in Section 3. The detailed proofs of the theorems are given in Appendix.

2 Main results

For a column vector α, let $\alpha^{\otimes 0} = 1$, $\alpha^{\otimes 1} = \alpha$, and $\alpha^{\otimes 2} = \alpha\alpha^{\mathrm{T}}$. For a matrix A or a vector a, $\|A\| = \sup_{i,j} |a_{ij}|$ and $\|a\| = \sup_i |a_i|$. We introduce the following notations:

$$
\begin{aligned}
&S^{(d)}(\beta,t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t) Z_i(t)^{\otimes d} e^{\beta^{\mathrm{T}} Z_i(t)}, \quad s^{(d)}(\beta,t) = \mathbb{E}S^{(d)}(\beta,t), \quad (d=0,1,2),\\
&\bar{Z}(\beta,t) = \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}, \qquad \mu(\beta,t) = \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)},\\
&V(\beta,t) = \frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)} - \bar{Z}(\beta,t)^{\otimes 2}, \quad v(\beta,t) = \frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)} - \mu(\beta,t)^{\otimes 2},\\
&S^{(d)*}(\beta,t) = \frac{1}{n}\sum_{i=1}^{n} w_i Y_i(t) Z_i(t)^{\otimes d} e^{\beta^{\mathrm{T}} Z_i(t)},\\
&\bar{Z}^*(\beta,t) = \frac{S^{(1)*}(\beta,t)}{S^{(0)*}(\beta,t)}, \quad V^*(\beta,t) = \frac{S^{(2)*}(\beta,t)}{S^{(0)*}(\beta,t)} - \bar{Z}^*(\beta,t)^{\otimes 2},
\end{aligned}
$$

where "$\mathbb{E}$" stands for the expectation.

For counting processes $(N_i(t), t \in [0,\tau])$, we have the following Doob-Meyer decomposition:

$$N_i(t) = M_i(t) + \int_0^t Y_i(u)\lambda_0(u) e^{\beta_0^{\mathrm{T}} Z_i(u)} du,$$

where β_0 is the true parameter, $\lambda_0(t)$ is the baseline hazard function. $(M_i(t), t \in [0,\tau])$ is a martingale with respect to filtration $\{\mathcal{F}_{t,i}, t \in [0,\tau]\}$, and $\mathcal{F}_{t,i} = \sigma(N_i(s), Y_i(s), Z_i(s) : 0 \leqslant s \leqslant \tau)$, the σ-algebra generated by $\{N_i(s), Y_i(s), Z_i(s) : 0 \leqslant s \leqslant \tau\}$. They are square-integrable martingale sequences with

$$\langle M_i, M_i\rangle(t) = \int_0^t Y_i(u)e^{\beta_0^{\mathrm{T}} Z_i(u)}\lambda_0(u)du, \quad \langle M_i, M_j\rangle = 0, \quad i \neq j,$$

i.e. M_i and M_j are orthogonal when $i \neq j$. For a given n, both $M_i(t)$ and $\sum_{i=1}^n M_i(t)$ are martingales with respect to $\mathcal{F}_t^{(n)} = \bigcup_{i=1}^n \mathcal{F}_{t,i}$.

We need the following assumptions before stating our main results.

Assumption A (Noninformativeness). T_i is independent of the corresponding censoring variable conditionally on Z_i.

Assumption B (Finite interval). $\int_0^\tau \lambda_0(t)dt < \infty$ and $P(Y_i(t) = 1, \text{for all } t \in [0, \tau]) > 0$.

Assumption C (Asymptotic stability). For $d = 0, 1, 2$, there exists a neighborhood $\mathcal{B}$ of β_0 such that

$$\sup_{\beta\in\mathcal{B}, t\in[0,\tau]} \|S^{(d)}(\beta, t) - s^{(d)}(\beta, t)\| = o_p(1).$$

Assumption D (Covariates). $|Z_{ij}(0)| + \int_0^t |dZ_{ij}(t)| < C_Z$, a.s., for all i, j, and some constant $C_Z < \infty$.

Assumption E (Asymptotic regularity). For all $(\beta, t) \in \mathcal{B} \times [0, \tau]$,

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial\beta} s^{(0)}(\beta, t), \quad s^{(2)}(\beta, t) = \frac{\partial^2}{\partial\beta^2} s^{(0)}(\beta, t).$$

$s^{(d)}(\beta, t)$, $d = 0, 1, 2$, are continuous functions of $\beta \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$. $s^{(0)}(\beta, t)$ is also bounded away from 0 on $\mathcal{B} \times [0, \tau]$. The matrix

$$\Sigma = \int_0^\tau v(\beta_0, t)s^{(0)}(\beta_0, t)\lambda_0(t)dt$$

is positive definite.

Assumption F (Weighting variable). $w_1, w_2, \ldots$ are i.i.d. with $P(w_1 > 0) = 1$, $E(w_1) = 1$, and $Ew_1^2 = \pi \geqslant 1$. The sequences $\{w_i\}$ and $\{T_i, \delta_i, Z_i\}$ are independent.

Note that the partial derivative conditions on $s^{(d)}$ are also satisfied by $S^{(d)}(d = 0, 1, 2)$. It was shown in [4] that under the conditions similar to Assumptions A–E,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma^{-1}), \tag{2.1}$$

where "$\xrightarrow{\mathcal{L}}$" stands for convergence in distribution.

In the rest of this paper, notations $\mathcal{L}^*$, P^* denote the corresponding probability calculations conditionally on $\{T_i, \delta_i, \mathbf{Z}_i\}_{i=1}^n$. Our main results are as follows.

Theorem 1. *Suppose that Assumptions A–F hold. We have*

$$\sqrt{n}(\hat{\beta}_n^* - \beta_0) = \Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n w_i \int_0^\tau (Z_i(t) - \mu(\beta_0, t))dM_i(t) + o_p(1). \tag{2.2}$$

In particular, if $w_i \equiv 1$, we have

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau (Z_i(t) - \mu(\beta_0, t))dM_i(t) + o_p(1). \tag{2.3}$$

Obviously, it follows from (2.3) that we have the well-known result (2.1).

Theorem 2. *Suppose that Assumptions A–F hold and* $\pi = 2$. *Then*

$$\mathcal{L}^*(\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n)) \to N(0, \Sigma^{-1}) \quad \textit{in probability}, \tag{2.4}$$

as $n \to \infty$. *Furthermore,*

$$\sup_{\mathbf{u} \in R^p} |P^*(\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n) \leqslant \mathbf{u}) - P(\sqrt{n}(\hat{\beta}_n - \beta_0) \leqslant \mathbf{u})| \to 0, \quad \textit{in probability}, \tag{2.5}$$

as $n \to \infty$, *where* $\mathbf{u}$ *runs over all* p*-vectors, and the inequality between vectors means coordinate-wise inequality.*

Note that when $\pi = 1$ the random weighting variables are constant 1, and the random weighting estimate $\hat{\beta}_n^* = \hat{\beta}_n$, a.s., the result of Theorem 1 reduces to the classical result of weak consistency of maximum partial likelihood estimates. The requirement of $E(w_i^2) = 2$ in Theorem 2 is to ensure the equality of the limiting variances of the random weighting estimator and the original maximum partial likelihood estimator. Examples of distributions for the weighting variables satisfying this condition include the exponential distribution with mean 1 and Poisson distribution with mean 1, among others.

Theorem 2 has an important implication that, for a large sample size, the distribution of the maximum partial likelihood estimator of β can be approximated by that of random weighting estimates. In practical applications, this can be done by Monte Carlo method. Specifically, one can generate random weights repeatedly for (1.2) and then solve the respective estimates of the regression parameters. Then the empirical distribution of the produced estimates is used as an approximation to the distribution of the maximum partial likelihood estimator of β. For example, in deriving the $(1 - \alpha)100\%$ confidence interval for the partial likelihood estimate, $\hat{\beta}_n$, from a particular data set, one can implement the random weighting m times to obtain the estimates $\hat{\beta}_n^{*(1)}, \hat{\beta}_n^{*(2)}, \ldots, \hat{\beta}_n^{*(m)}$ and hence use the lower and upper $\alpha/2$ quantiles of these quantities as the approximation of lower and upper limits of the confidence interval. The results in Theorem 2 validate this procedure.

3 Simulation studies

We conduct simulation studies to evaluate the performance of the random weighting method for the Cox model. Survival data are generated from an exponential regression model with hazard rate $\lambda(t|X = x) = \lambda_0(t)\exp(\beta x)$ and baseline hazard rate $\lambda_0(t) = 1$. We take $\beta = 0$. Covariates are taken to be 0 and 1's with one-half for each. The random weighting method is compared with bootstrap method. For the random weighting method, we generate i.i.d. w_i in (1.2) from the exponential distribution (the exponential random weighting method) and Poisson distribution (the Poisson random weighting method) with means 1 respectively. Two bootstrap methods are compared with the random weighting method, namely, the case bootstrap method and the conditional bootstrap method (see Sections 3.5 and 7.3 in [6] for details). These two methods are frequently used in survival analysis. The former one is the well known resampling method which simply resamples observations with replacement, while the latter simulates failure times based on the estimated survival function.

We randomly generate 50 survival times using the censoring rates 36.8%, 60.7% and 77.9% respectively. Then we estimate β by the exponential random weighting method, the Poisson random weighting method, the case bootstrap method and the conditional bootstrap method. Simulation replicates is 500. Histograms of the random weighted estimates are given in Figures 1–3. From Figures 1–2, it can been seen that all histograms appear to be bell shaped. But for censoring rate 77.9% (heavey censoring), only the histogram by the exponential random weighting method appears to be bell shaped and other histograms are skewed in Figure 3. This is because in the case of heavy censoring and small sample size, the censored data may be repeatedly resampled and the information in failure times is not enough for producing a good estimate of β. On the other hand, the Poisson random weighting method produces an average 0 weight of proportion 36.8% in $\{w_i\}$. If there are too many zero weights on failure times, similarly to bootstrap methods, the weighted partial likelihood equation with i.i.d. Poisson weights does not produce a good estimate of β.

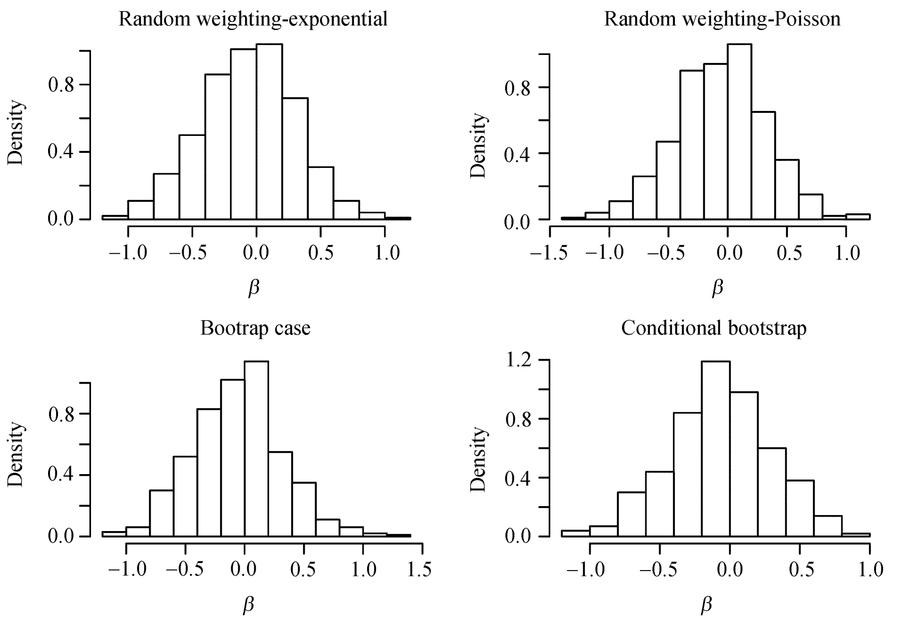

Figure 1 Histogram of the estimate of β by using the random weighting and bootstrap methods with sample size $n = 50$ and 500 replications ($\hat{\beta} = -0.0635$, censoring rate= 36.8%)

To compare the random weighting method with bootstrap methods in confidence interval estimation, we construct the confidence intervals based on the exponential random weighting method (RW-exp), the Poisson random weighting method (RW-Poission), the asymptotic normal distribution of maximum partial likelihood estimates (Asy-normal), the conditional bootstrap method (BS-cond) and the case bootstrap method (BS-case). The sample sizes are set to be 20, 40, 100, 200, 400, and 1,000 respectively, and as before, the number of replicates is 500. Simulation results are given in Tables 1 for censoring rates 0.2 and 0.5. From Table 1, it can be seen that the empirical coverage rates are reasonably close to the nominal ones for all methods in all cases excluding the case with sample size 20 and censoring rate 0.5. Not

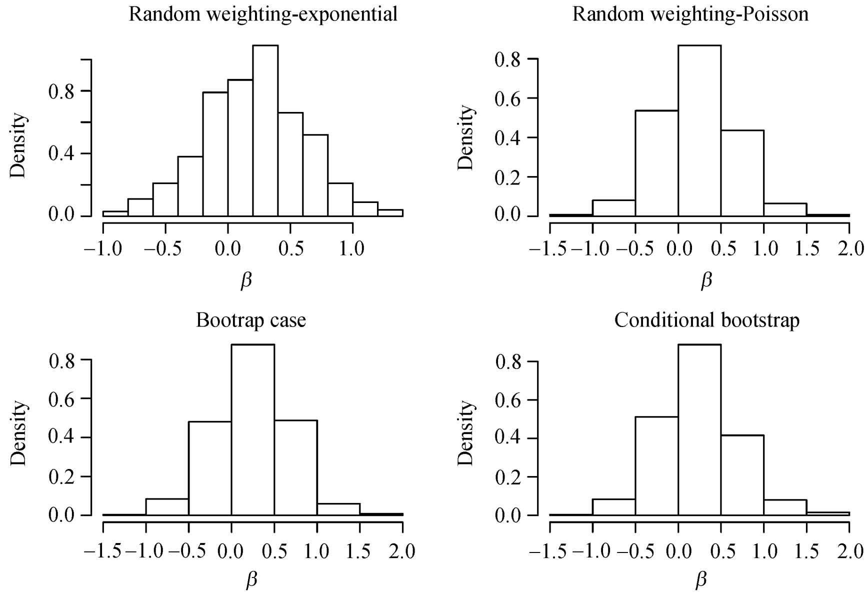

Figure 2 Histogram of the estimate of β by using the random weighting and bootstrap methods with sample size $n = 50$ and 500 replications ($\hat{\beta} = 0.2162$, censoring rate= 60.7%)

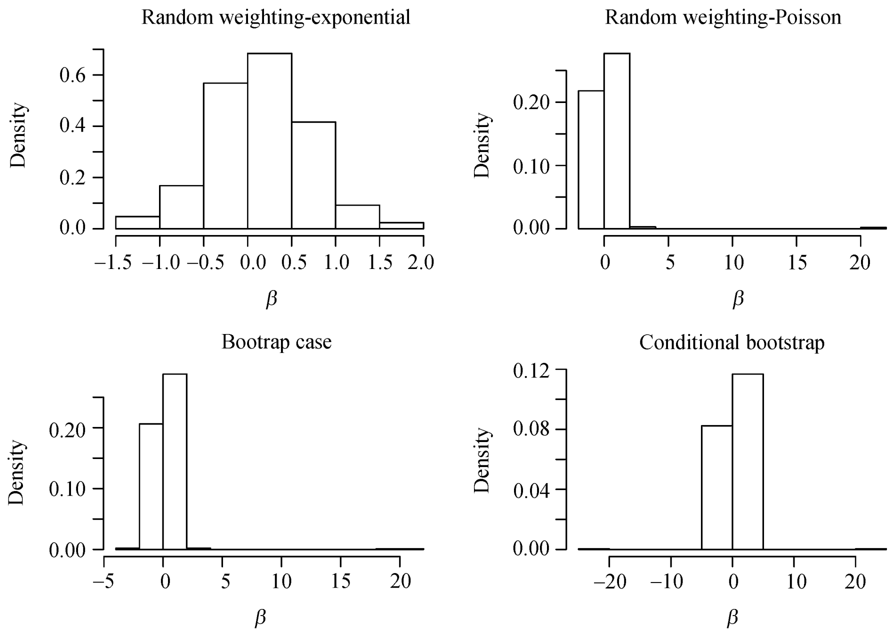

Figure 3 Histogram of the estimate of β by using the random weighting and bootstrap methods with sample size $n = 50$ and 500 replications ($\hat{\beta} = 0.1399$, censoring rate= 77.9%)

unexpectedly, the length of confidence intervals decreases with sample sizes and increases with the censoring rates. When the sample size is 40 and censoring rate is $\rho = 0.2$ or 0.5, the exponential random weighting method produces the shortest confidence interval. It can also be seen that the performance of the conditional bootstrap method is slightly better than that of the case bootstrap method, though the difference is not significant. The Poisson random

Table 1 Length of confidence intervals and empirical coverage rates (true parameter $\beta = 0$)

$1-\alpha$	n	ρ	Asy-normal	BS-cond	BS-case	RW-Exponential	RW-Poisson
95%	20	0.2	2.024 (0.946)	2.627 (0.920)	2.800 (0.930)	1.967 (0.905)	3.091 (0.938)
		0.5	——	——	——	2.638 (0.906)	——
	40	0.2	1.407 (0.951)	1.474 (0.937)	1.498 (0.938)	1.372 (0.933)	1.494 (0.945)
		0.5	1.801 (0.962)	1.963 (0.943)	2.004 (0.946)	1.769 (0.934)	2.003 (0.946)
	100	0.2	0.882 (0.947)	0.890 (0.940)	0.895 (0.940)	0.866 (0.936)	0.891 (0.939)
		0.5	1.119 (0.946)	1.133 (0.937)	1.139 (0.939)	1.102 (0.938)	1.135 (0.940)
	200	0.2	0.621 (0.958)	0.621 (0.951)	0.622 (0.952)	0.614 (0.955)	0.620 (0.956)
		0.5	0.788 (0.949)	0.789 (0.944)	0.792 (0.942)	0.778 (0.942)	0.791 (0.947)
	400	0.2	0.439 (0.947)	0.437 (0.942)	0.437 (0.941)	0.434 (0.942)	0.436 (0.942)
		0.5	0.556 (0.952)	0.553 (0.947)	0.554 (0.948)	0.550 (0.949)	0.552 (0.945)
	1000	0.2	0.277 (0.955)	0.275 (0.952)	0.275 (0.952)	0.274 (0.950)	0.274 (0.948)
		0.5	0.351 (0.950)	0.348 (0.948)	0.349 (0.948)	0.348 (0.947)	0.348 (0.948)
99%	20	0.2	2.660 (0.991)	4.689 (0.971)	5.418 (0.976)	2.580 (0.961)	8.763 (0.983)
		0.5	——	——	——	3.509 (0.967)	——
	40	0.2	1.849 (0.990)	1.957 (0.978)	2.001 (0.980)	1.790 (0.979)	2.018 (0.987)
		0.5	2.367 (0.997)	2.821 (0.987)	2.970 (0.988)	2.313 (0.981)	3.241 (0.990)
	100	0.2	1.159 (0.988)	1.158 (0.982)	1.164 (0.982)	1.122 (0.982)	1.165 (0.983)
		0.5	1.471 (0.993)	1.479 (0.983)	1.491 (0.984)	1.430 (0.984)	1.489 (0.986)
	200	0.2	0.817 (0.990)	0.804 (0.986)	0.807 (0.985)	0.795 (0.985)	0.804 (0.986)
		0.5	1.036 (0.991)	1.023 (0.989)	1.028 (0.982)	1.007 (0.988)	1.029 (0.987)
	400	0.2	0.576 (0.990)	0.566 (0.986)	0.565 (0.985)	0.561 (0.987)	0.565 (0.986)
		0.5	0.730 (0.991)	0.715 (0.986)	0.719 (0.987)	0.712 (0.987)	0.716 (0.987)
	1000	0.2	0.364 (0.990)	0.355 (0.985)	0.356 (0.986)	0.355 (0.986)	0.356 (0.984)
		0.5	0.461 (0.991)	0.450 (0.987)	0.452 (0.989)	0.450 (0.986)	0.450 (0.988)

$1-\alpha$ is the nominal coverage rate, ρ is the censoring rate. Data presentation: length of confidence interval (empirical coverage rate).

weighting method has a similar performance as the bootstrap methods. It is noted that the confidence interval constructed based on the asymptotic normal distribution of the maximum partial likelihood estimator is only slightly worse than the one constructed based on the exponential random weighting method for the case of moderate sample sizes and low censoring rate.

Consistent with the previous findings from Figures 1–3, by Table 1, the exponential random weighting method produces more reliable results. In Table 1, some of the cells are left blank if the corresponding confidence interval cannot be properly constructed. For a high censoring rate ($\rho = 0.5$) and a small sample size $n = 20$, only the exponential random weighting method can be used to construct a confidence interval. This is because with such a high censoring rate and the small sample size, the resampled data may be censored very heavily and does not have enough event-time information for obtaining a reasonable estimate of β.

4 Discussion

Bootstrap methods are widely used to approximate the empirical distribution of an estimator. They are especially useful if there is no explicit variance estimate or the variance estimate is not good in case that sample size is not large enough. The random weighting method is similar to bootstrap method, which can also be used to approximate the empirical distribution of the estimator. The main difference is that random weighting method does not resample the observations with replacement. This characteristic of the random weighting method makes it more appealing than other resampling methods when the sample size is small and the censoring rate is high. This is because the resampling method can easily result in a data set with a high censoring rate (consequently, the statistical inference becomes difficult) when the censoring rate in the original data is relatively high.

In this paper, a random weighting method is proposed for analyzing event-time data when a proportional hazard model is adopted. The consistency and asymptotic normality of the regression coefficient estimator by the random weighting method are proved. We conduct the simulation studies to investigate the properties of the proposed method and to compare it with bootstrap methods. It was shown that the performance of the random weighting method is generally similar to that of bootstrap methods but it is more favorable when the sample size is small and the censoring rate is high. The simulation studies are also carried out for comparing two choices of the weights in (1.2). All simulation studies confirm that the exponential random weighing method has been performed superiorly over other methods.

Acknowledgements The authors would like to thank the two anonymous referees for their constructive suggestions and comments.

References

1 Cox D R. Regression models and lifetables (with discussion). *J Roy Statist Soc Ser B*, **34**: 187–220 (1972)
2 Cox D R. Partial likelihood. *Biometrika*, **62**: 269–276 (1975)
3 Ansersen P K, Gill R D. Cox's regression model for counting processes: a large sample study. *Ann Statist*, **10**: 1100–1120 (1982)
4 Fleming T R, Harrington D P. Counting Processes and Survival Analysis. New York: John Wiley & Sons, Inc., 1991
5 Efron B. Censored data and the bootstrap. *J Amer Statist Assoc*, **76**: 312–319 (1981)
6 Davison A C, Hinkley D V. Bootstrap Methods and Their Application. Cambridge: Cambridge University Press, 1997
7 James L F. A study of a class of weighted bootstraps for censored data. *Ann Statist*, **25**: 1595–1621 (1997)
8 Efron B. Nonparametric standard errors and confidence intervals (with discussion). *Canad J Statist*, **9**: 139–319 (1981)
9 Hjort N L. Bootstrapping Cox's regression model. Technical report NSF-241, Department of Statistics, Stanford University, 1985
10 Zheng Z. Random weighting method (In Chinese). *Acta Math Appl Sin*, **10**: 247–253 (1987)
11 Rao C R, Zhao L C. Approximation to the distribution of M-estimaties in linear models by randomly weighted bootsrap. *Sankhyā*, **54**: 323–331 (1992)
12 Rubin D B. The Bayesian bootstrap. *Ann Statist*, **9**: 130–134 (1981)
13 Weng C S. On a second-order asymptotic property of the Bayesian bootstrap. *Ann Statist*, **17**: 705–710 (1989)
14 Kim Y, Lee J. Bayesian bootstrap for proportional hazards models. *Ann Statist*, **31**: 1905–1922 (2003)
15 Ortega J M, Rheinboldt W G. Iterative Solutions of Nonlinear Equations in Several Variable. New York: Academic Press, 1970
16 van der Vaart A W, Wellner J A. Weak Convergence and Empirical Processes. New York: Springer-Verlag, 1996

Appendix

For proving the theorems, we firstly give the following lemmas.

Lemma A1. *Under Assumptions A–E holding, if $\beta_n^* \in \mathcal{B}$ converges to β_0 in probability as $n \to \infty$, then we have*

$$\begin{aligned}
&\sup_{t\in[0,\tau]} \|S^{(d)*}(\beta_n^*, t) - s^{(d)}(\beta_0, t)\| = o_p(1), \quad d = 0, 1, 2,\\
&\sup_{t\in[0,\tau]} \|\bar{\mathbf{Z}}^*(\beta_n^*, t) - \mu(\beta_0, t)\| = o_p(1),\\
&\sup_{t\in[0,\tau]} \|V^*(\beta_n^*, t) - v(\beta_0, t)\| = o_p(1).
\end{aligned}$$

The proof of Lemma A1 is rather standard and therefore is omitted.

Lemma A2. *Suppose that Assumptions A–F hold, then, with probability one there exists a unique solution to* (1.2), $\hat{\beta}_n^*$, *such that*

$$\hat{\beta}_n^* \to \beta_0, \quad \textit{inprobability}.$$

Proof. Let $f_n(\beta) = \frac{1}{\sqrt{n}} U^*(\beta)$, where β satisfies

$$\|\beta - \beta_0\| = \rho_n, \text{ with } \rho_n \to 0 \text{ and } \sqrt{n}\rho_n^2 \to \infty \text{ as } n \to \infty. \tag{A.1}$$

By Taylor expansion, there exists β^* which is in the line segment of β and β_0, such that

$$f_n(\beta) = f_n(\beta_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \int_0^\tau V^*(\beta^*, t) dN_i(t)(\beta - \beta_0). \tag{A.2}$$

By Lemma A1 and $\frac{1}{n}\sum_{i=1}^n w_i \int_0^\tau v(\beta_0, t) dN_i(t) = O_p(1)$, it can be shown that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n w_i \int_0^\tau V^*(\beta^*, t) dN_i(t) &= \frac{1}{n}\sum_{i=1}^n w_i \int_0^\tau v(\beta_0, t) dN_i(t) + o_p(1)\\
&\xrightarrow{P} \mathbb{E}\bigg(w_1 \int_0^\tau v(\beta_0, t) dN_1(t)\bigg) = \mathbb{E}\bigg(\int_0^\tau v(\beta_0, t) Y_1(t) e^{\beta_0^{\mathrm{T}} Z_1(t)} \lambda_0(t) d(t)\bigg) = \Sigma,
\end{aligned} \tag{A.3}$$

where "$\xrightarrow{P}$" stands for the convergence in probability. By (A.1), (A.2), (A.3) and Assumption E, it follows that

$$\sup_{\|\beta-\beta_0\|=\rho_n} (\beta - \beta_0)^{\mathrm{T}} f_n(\beta) \leqslant O_p(\rho_n) - (\lambda_{\min}(\Sigma) + o_p(1))\sqrt{n}\rho_n^2 \xrightarrow{P} -\infty,$$

where $\lambda_{\min}(\Sigma)$ is the minimum eigenvalue of Σ. Therefore,

$$\lim_{n\to\infty} P\bigg(\sup_{\|\beta-\beta_0\|=\rho_n} (\beta - \beta_0)^{\mathrm{T}} f_n(\beta) < 0\bigg) = 1.$$

Therefore, based on the Leray-Schauder Theorem and its corollary[15], it follows that with probability one there exists a unique solution to (1.2), $\hat{\beta}_n^*$, which converges to β_0 in probability.

Proof of Theorem 1. By Taylor expansion, there exists β^*, which is in the line segment of $\hat{\beta}_n^*$ and β_0, such that

$$\begin{aligned}\frac{1}{\sqrt{n}}U^*(\beta_0) &= \Big[-\frac{1}{n}\frac{\partial U^*(\beta)}{\partial \beta^{\mathrm{T}}}\Big|_{\beta=\beta^*}\Big]\sqrt{n}(\hat{\beta}_n^* - \beta_0) \\ &= \Big[\frac{1}{n}\sum_{i=1}^n w_i \int_0^\tau V^*(\beta^*, t)dN_i(t)\Big]\sqrt{n}(\hat{\beta}_n^* - \beta_0), \qquad \text{(A.4)}\end{aligned}$$

since $U^*(\hat{\beta}_n^*) = 0$ and $\frac{\partial U^*(\beta)}{\partial \beta^{\mathrm{T}}} = \sum_{i=1}^n w_i \int_0^\tau V^*(\beta, t)dN_i(t)$. By Lemma A1 together with the Doob-Meyer decomposition for $N_i(t)$ and the consistency of $\hat{\beta}_n^*$, we have

$$\frac{1}{n}\sum_{i=1}^n w_i \int_0^\tau v(\beta_0, t)dN_i(t) \xrightarrow{P} \mathbb{E}\Big(w_1 \int_0^\tau v(\beta_0, t)Y_1(t)e^{\beta_0^{\mathrm{T}} Z_1(t)}\lambda_0(t)dt\Big) = \Sigma \qquad \text{(A.5)}$$

and

$$\begin{aligned}&\Big\|\frac{1}{n}\sum_{i=1}^n w_i \int_0^\tau (V^*(\beta^*, t) - v(\beta_0, t))dN_i(t)\Big\| \\ &\qquad \leqslant \sup_{t\in[0,\tau]} \|V^*(\beta^*, t) - v(\beta_0, t)\| \frac{1}{n}\sum_{i=1}^n w_i N_i(\tau) = o_p(1). \qquad \text{(A.6)}\end{aligned}$$

Therefore in view of (A.5) and (A.6) it follows that

$$\frac{1}{n}\sum_{i=1}^n w_i \int_0^\tau V^*(\beta^*, t)dN_i(t) = \Sigma + o_p(1). \qquad \text{(A.7)}$$

By (A.4) and (A.7), we have

$$\sqrt{n}(\hat{\beta}_n^* - \beta_0) = (\Sigma + o_p(1))^{-1}\frac{1}{\sqrt{n}}U^*(\beta_0) = (\Sigma^{-1} + o_p(1))\frac{1}{\sqrt{n}}U^*(\beta_0). \qquad \text{(A.8)}$$

Divide $\frac{1}{\sqrt{n}}U^*(\beta_0)$ into two parts as follows:

$$\begin{aligned}\frac{1}{\sqrt{n}}U^*(\beta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i \int_0^\tau (Z_i(t) - \mu(\beta_0, t))dM_i(t) \\ &\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i \int_0^\tau (\bar{Z}^*(\beta_0, t) - \mu(\beta_0, t))dM_i(t). \qquad \text{(A.9)}\end{aligned}$$

Since $\{w_i \int_0^\tau (Z_i - \mu(\beta_0, t))dM_i(t)\}_{n\geqslant 1}$ are i.i.d. with zero mean vector and

$$\begin{aligned}\mathrm{Var}\Big(w_1 \int_0^\tau (Z_1(t) - \mu(\beta_0, t))dM_1(t)\Big) &= \pi \int_0^\tau \mathbb{E}(Z_1(t) - \mu(\beta_0, t))^{\otimes 2} Y_1(t)e^{\beta_0^{\mathrm{T}} Z_1(t)}\lambda_0(t)d(t) \\ &= \pi\Sigma,\end{aligned}$$

by multivariate central limit theorem, it follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n w_i \int_0^\tau (Z_i(t) - \mu(\beta_0, t))dM_i(t) \xrightarrow{\mathcal{L}} N(0, \pi\Sigma).$$

Note that, conditionally on $w_1, \ldots, w_n$, $w_i \int_0^t (\bar{Z}^*(\beta_0, u) - \mu(\beta_0, u))dM_i(u)$ is a martingale with respect to $\{\mathcal{F}_t, 0 \leqslant t \leqslant \tau\}$ with zero mean vector, which means that the expectation of the second term of the right-hand side of (A.9) is zero vector. By Lemma A1, together with the martingale property and $\langle M_i, M_j \rangle = 0\,(i \neq j)$, we have

$$\left\| \mathbb{E}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \int_0^\tau (\bar{Z}^*(\beta_0, t) - \mu(\beta_0, t))dM_i(t) \right]^{\otimes 2} \right\| = o(1).$$

Thus, the second term of the right-hand side of (A.9) is asymptotically neglectable in probability. Hence, in view of (A.9), it is obvious that

$$\frac{1}{\sqrt{n}} U^*(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \int_0^\tau (Z_i(t) - \mu(\beta_0, t))dM_i(t) + o_p(1) \xrightarrow{\mathcal{L}} N(0, \pi\Sigma). \tag{A.10}$$

Thus, by (A.8) and (A.10), we have (2.2) since

$$\sqrt{n}(\hat{\beta}_n^* - \beta_0) = (\Sigma^{-1} + o_p(1)) \frac{1}{\sqrt{n}} U^*(\beta_0) = \Sigma^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \int_0^\tau (Z_i - \mu(\beta_0, t))dM_i(t) + o_p(1).$$

Proof of Theorem 2. By the result of Theorem 1, we have

$$\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n) = \Sigma^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (w_i - 1) \int_0^\tau (Z_i(t) - \mu(\beta_0, t))dM_i(t) + o_p(1). \tag{A.11}$$

From Lemma 2.9.5 of [16], it follows that conditionally on $\{T_i, \delta_i, Z_i(\cdot)\}_{i \geqslant 1}$,

$$\mathcal{L}^* \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (w_i - 1) \int_0^\tau (Z_i(t) - \mu(\beta_0, t))dM_i(t) \right) \to N(0, \Sigma) \tag{A.12}$$

for almost every sequence $\{T_i, \delta_i, Z_i(\cdot)\}_{i \geqslant 1}$. Thus, by (A.11) and (A.12), it can be shown that

$$\mathcal{L}^*(\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n)) \to N(0, \Sigma^{-1}) \quad \text{in probability.}$$

Similar to the discussion in [11], (2.5) can be shown to hold true. The proof is thus completed.