Lifetime Data Anal (2007) 13:399–420
DOI 10.1007/s10985-007-9043-3

Maximum likelihood estimation for tied survival data under Cox regression model via EM-algorithm

Thomas H. Scheike · Yanqing Sun

Received: 23 October 2006 / Accepted: 22 June 2007 / Published online: 8 August 2007

Abstract We consider tied survival data based on Cox proportional regression model. The standard approaches are the Breslow and Efron approximations and various so called exact methods. All these methods lead to biased estimates when the true underlying model is in fact a Cox model. In this paper we review the methods and suggest a new method based on the missing-data principle using EM-algorithm that leads to a score equation that can be solved directly. This score has mean zero. We also show that all the considered methods have the same asymptotic properties and that there is no loss of asymptotic efficiency when the tie sizes are bounded or even converge to infinity at a given rate. A simulation study is conducted to compare the finite sample properties of the methods.

Keywords Cox regression model · Tied survival data · EM-algorithm · Asymptotics

1 Introduction

Assume that we have n independent right-censored survival data that follow Cox regression model (Cox 1972). Formally, assume that, for $1 \leq i \leq n$, the survival times X_i and censoring times C_i are conditionally independent given a p-dimensional covariate Z_i, and that we observe $U_i = \min(X_i, C_i)$ as well as the censoring indicator

T. H. Scheike
Department of Biostatistics, University of Copenhagen, Øster Farimagsgade 5 B, P.O.B. 2099,
1014 Copenhagen K, Denmark
e-mail: ts@biostat.ku.dk

Y. Sun (✉)
Department of Mathematics and Statistics, The University of North Carolina at Charlotte,
9201 University City Boulevard, Charlotte, NC, 28223, USA
e-mail: yasun@uncc.edu

$\delta_i = I(X_i \le C_i)$. We assume that given Z_i the intensity for X_i is given by

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp(Z_i^T\beta)$$

where $Y_i(t) = I(U_i \ge t)$ is the at-risk indicator, $\lambda_0(t)$ is the non-parametric baseline, and β is the p-dimensional regression coefficients.

Observations based on Cox regression model will often be subject to additional coarsening that leads to tied survival data, and the objective in this paper is to review the existing methods and suggest a new method that leads to unbiased estimates. In addition we point out that the tied survival data leads to large sample properties equivalent to those based the underlying un-tied survival data. In practice ties are often observed and it is therefore crucial to know how to deal with these.

The standard approaches are the Breslow (1972) and Efron (1977) approximations that are simple to implement; see also Therneau and Grambsch (2000). Let T_1 and T_2 be observed tied survival ($T_1 = T_2$) data based on Cox regression model with two observations tied at this value. Define $R_i(t) = Y_i(t)\exp(Z_i^T\beta)$. If the data were untied the partial likelihood contribution from T_1 and T_2 would be either

$$\frac{R_1(T_1)}{R_1(T_1) + R_2(T_1) + \cdots + R_n(T_1)}\,\frac{R_2(T_2)}{R_2(T_2) + R_3(T_2) + \cdots + R_n(T_2)}$$

if T_1 actually came before T_2 or

$$\frac{R_2(T_2)}{R_1(T_2) + R_2(T_2) + \cdots + R_n(T_2)}\,\frac{R_1(T_1)}{R_1(T_1) + R_3(T_1) + \cdots + R_n(T_1)}$$

if T_2 actually came before T_1.

The Breslow approximation (Breslow 1972; Peto 1972) uses $\sum_i R_i(T_2)$ in both denominators and thus uses the approximation

$$\frac{R_1(T_1)}{R_1(T_2) + R_2(T_2) + \cdots + R_n(T_2)}\,\frac{R_2(T_2)}{R_1(T_2) + R_2(T_2) + \cdots + R_n(T_2)}.$$

The Efron approximation (Efron 1977), in contrast uses the approximation

$$\frac{R_1(T_1)}{R_1(T_2) + R_2(T_2) + \cdots + R_n(T_2)}\,\frac{R_2(T_2)}{0.5(R_1(T_2) + R_2(T_2)) + \cdots + R_n(T_2)}$$

and thus takes an average of the two relative-risk terms $R_1(T_2)$ and $R_2(T_2)$.

With ties of size k ($T_1 = T_2 = \cdots = T_k$) the Breslow approximation will use

$$\prod_{i=1}^{k} \frac{R_i(T_1)}{R_1(T_1) + R_2(T_1) + \cdots + R_n(T_1)}$$

and the Efron approximation becomes

$$\prod_{i=1}^{k} \frac{R_i(T_1)}{\frac{k-i+1}{k}\sum_{j=1}^{k} R_j(T_1) + R_{k+1}(T_1) + \cdots + R_n(T_1)}.$$

Both suggestions will result in score functions whose expectations are not equal to 0, and therefore will lead to biased estimates. The Breslow estimator leads to estimates that are shrunk towards 0. As we will point out later, even though both methods lead to biased estimates their asymptotic performance are, however, equivalent to that of the score based on fully observed un-tied data.

There are several so called exact solutions that involve more extensive computations, but these methods are also ad-hoc and do not appear to improve on the Efron approximation. We omit the details of various exact procedures and refer to Therneau and Grambsch (2000) and Kalbfleisch and Prentice (2002). Different statistical softwares may use different exact procedures. As far as we know, there have been very little study on the validity and efficiency of the various approximation methods for tied survival data under Cox regression model. A previous paper by Hertz-Picciotto and Rockhill (1997) examining this question with simulations considered only two sample case without censoring.

In this paper, we suggest an alternative procedure based on the EM-algorithm that is easy to implement and is fully efficient. Our approach is related to the rank-based techniques for interval censored data as described in Satten (1996), but is considerably simpler to implement because we only consider permutations within each tie. Our procedure based on the EM-algorithm is described in Sect. 2. The asymptotic properties of the estimators using the EM-algorithm, the Breslow approximation and Efron approximation are derived in Sect. 3. An extensive simulation study is done in Sect. 4 to evaluate the EM-algorithm and to compare it with the existing methods.

2 An EM procedure for tied survival data

Let $T_1, \ldots, T_J$ be distinct ordered and possibly censored survival times with n_j ties at time T_j. The covariates associated with time T_j are $Z_{j,k}, k = 1, \ldots, n_j$. Let $n = \sum_{j=1}^{J} n_j$. We assume that the underlying survival times $X_i, i = 1, .., n$, arise from the Cox regression model as described in the beginning of Sect. 1. For each j, $j = 1, \ldots, J$, let $\{T^*_{j,k}, k = 1, \ldots, n_j\}$ be the survival/censoring times tied at T_j, with $T^*_{j,k} \leq T^*_{j,m}$ for $k \leq m$. We assume that any two tie clusters are correctly ordered such that for any $j < l$: $\max_k(T^*_{j,k}) < \min_k(T^*_{l,k})$. For simplicity we also assume that each tie cluster consists of either observed survival times or censoring times only. In the situation where censoring times are tied with observed survival times, one can split the tied censoring and observed survival times into two tie clusters and place the tied censoring times right before the tied observed survival times. Given the observed tied survival times, it is unknown how the covariates $\{Z_{j,k}, k = 1, \ldots, n_j\}$ relate to $\{T^*_{j,k}, k = 1, \ldots, n_j\}$. We propose an EM-algorithm that deals with this situation to estimate β as well as the cumulative baseline function of the Cox model.

For fixed j let $P(n_j)$ denote the set of all permutations of the n_j indexes $\{1, 2, \ldots, n_j\}$ at time T_j, and let $p = \{i_1, i_2, \ldots, i_{n_j}\}$ be a permutation of the indexes. The true ordering of the covariates $\{Z_{j,k}, k = 1, \ldots, n_j\}$ that relate to $\{T^*_{j,k}, k = 1, \ldots, n_j\}$ is denoted by the random vector $P_j = \{I_1, I_2, \ldots, I_{n_j}\}$ and is unobserved.

Let $N^*_{j,k}(t)$ be the counting process and $Y^*_{j,k}(t)$ the at risk indicator associated with the survival time $T^*_{j,k}$ and its censoring indicator. The full data is the situation where we know which covariates correspond to $T^*_{j,k}$, thus leading to the triplets

$$(N^*_{j,k}(t), Y^*_{j,k}(t), Z_{j,I_k}), \quad j = 1, \ldots, J, \quad k = 1, \ldots, n_j,$$

where Z_{j,I_k} is the covariate related to the (j, k)th counting process.

We pretend to observe $\{(N^*_{j,k}(t), Y^*_{j,k}(t)),\ \ j = 1, \ldots, J, k = 1, \ldots, n_j\}$ and $\{Z_{j,k},\ j = 1, \ldots, J, k = 1, \ldots, n_j\}$ where the first index j is for the jth distinct tie, but we do not know how the second index for (N^*, Y^*) and Z are related. That is, we do not know which one of $Z_{j,1}, \ldots, Z_{j,n_j}$ is the covariate for $(N^*_{j,k}, Y^*_{j,k})$. We denote this data as D^*. Later as we shall see the obtained score will not depend on the values $T^*_{j,k}$, and therefore will also be an efficient score when these are not observed.

With the full data the likelihood can be written as (Andersen et al. 1993)

$$L = \prod_{j=1}^{J} \prod_{k=1}^{n_j} \prod_{t \le \tau} (Y^*_{j,k}(t) d\Lambda_0(t) \exp(Z^T_{j,I_k}\beta))^{dN^*_{j,k}(t)}$$
$$\exp\left(-\int_0^\tau Y^*_{j,k}(t) \exp(Z^T_{j,I_k}\beta) d\Lambda_0(t)\right),$$

and the log-likelihood is

$$l = \sum_{j=1}^{J} \sum_{p \in P(n_j)} I(p = P_j) \sum_{k=1}^{n_j} \left\{ \int_0^\tau (\log(d\Lambda_0(t)) + Z^T_{j,i_k}\beta) dN^*_{j,k}(t) \right.$$
$$\left. - \int_0^\tau Y^*_{j,k}(t) \exp(Z^T_{j,i_k}\beta) d\Lambda_0(t) \right\},$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds$ is the cumulative baseline function.

The expectation of this quantity given the extended version of the data D^* (where only the ordering of the covariates related to the survival times within each tie is unobserved) is

$$E(l|D^*) = \sum_{j=1}^{J} \sum_{p \in P(n_j)} E(I(p = P_j)|D^*) \tag{1}$$
$$\sum_{k=1}^{n_j} \left\{ \int_0^{\tau} (\log(d\Lambda_0(t)) + Z_{j,i_k}^T \beta) dN_{j,k}^*(t) \right.$$
$$\left. - \int_0^{\tau} Y_{j,k}^*(t) \exp(Z_{j,i_k}^T \beta) d\Lambda_0(t) \right\}.$$

Let $w_p(j, \beta) = E(I(p = P_j)|D^*, \beta)$. For uncensored survival times $T_{j,1}^*, \ldots, T_{j,n_j}^*$ that are included in D^* this expectation is nothing but the Cox partial likelihood for the data points in the jth tie cluster

$$w_p(j, \beta) = E(I(P_j = \{i_1, .., i_{n_j}\}|D^*, \beta) = \prod_{k=1}^{n_j} \frac{\exp(Z_{j,i_k}^T \beta)}{\sum_{l \geq k} \exp(Z_{j,i_l}^T \beta)}. \tag{2}$$

It is clear that $\sum_{p \in P(n_j)} w_p(j, \beta) = 1$. If the jth cluster consists of only censored survival times, then the weights $w_p(j, \beta)$ needs not to be calculated as it should become clear in the discussion following the score function (3). In the E-step of the EM-algorithm, we replace $E(l|D^*)$ in (1) by $w_p(j, \beta^m)$ where β^m is the previous iterative estimate of the parameter β. In the M-step that follows, we maximize $E(l|D^*)$ with respect to β and $\Lambda_0(\cdot)$ for the fixed weights $w_p(j, \beta^m),\ j = 1, \ldots, J$.

Define

$$S_v^{EM}(t, \beta, \beta^m) = \sum_{j=1}^{J} \sum_{p \in P(n_j)} w_p(j, \beta^m) \sum_{k=1}^{n_j} Z_{j,i_k}^{\otimes v} Y_{j,k}^*(t) \exp(Z_{j,i_k}^T \beta),$$

for $v = 0, 1, 2$, where $Z_{j,i_k}^{\otimes 0} = 1,\ Z_{j,i_k}^{\otimes 1} = Z_{j,i_k},\ Z_{j,i_k}^{\otimes 2} = Z_{j,i_k}(Z_{j,i_k})^T$.

The derivative $E(l|D^*)$ with respect to $d\Lambda_0(t)$ gives

$$\sum_{j=1}^{J} \sum_{p \in P(n_j)} w_p(j, \beta^m) \sum_{k=1}^{n_j} \left\{ \frac{dN_{j,k}^*(t)}{d\Lambda_0(t)} - Y_{j,k}^*(t) \exp(Z_{j,i_k}^T \beta) \right\}.$$

The root of the estimating function is solved as

$$d\tilde{\Lambda}_0(t) = \frac{dN_{\cdot,\cdot}^*(t)}{S_0^{EM}(t, \beta, \beta^m)},$$

where $N_{\cdot,\cdot}^*(t) = \sum_{j,k} N_{j,k}^*(t)$.

Taking the derivative of $E(l|D^*)$ with respect to β for the fixed weights $w_p(j, \beta^m)$ and plugging the above expression of $d\tilde{\Lambda}_0(\cdot)$ for $d\Lambda_0(\cdot)$, we obtain the following score function for β:

$$\begin{aligned}U_{EM}(\beta) &= \sum_{j=1}^{J} \sum_{p\in P(n_j)} w_p(j,\beta^m) \left\{ \sum_{k=1}^{n_j} \int_0^\tau Z_{j,i_k} dN^*_{j,k}(t) \right. \\ &\quad \left. - \int_0^\tau Z_{j,i_k} Y^*_{j,k}(t) \exp(Z^T_{j,i_k}\beta) d\tilde{\Lambda}_0(t) \right\} \\ &= \sum_{j=1}^{J} \sum_{p\in P(n_j)} w_p(j,\beta^m) \sum_{k=1}^{n_j} \int_0^\tau (Z_{j,i_k} - \frac{S_1^{EM}(t,\beta,\beta^m)}{S_0^{EM}(t,\beta,\beta^m)}) dN^*_{j,k}(t). \quad (3)\end{aligned}$$

The score (3) depends on the weights for censored clusters only through $S_v^{EM}(t,\beta,\beta^m)$ at uncensored survival times. At an uncensored survival time $t = T^*_{j,k}$ in the jth tie cluster, $\sum_{k=1}^{n'_j} Z^{\otimes v}_{j',i_k} Y^*_{j',k}(t) \exp(Z^T_{j',i_k}\beta)$ does not depend on permutations for $j' \neq j$ since the risk indicators for all subjects in other tie clusters are either 0 or 1. Since $\sum_{p\in P(n_j)} w_p(j,\beta^m) = 1$, the score (3) does not depend on the weights of clusters whose survival times are censored.

When the covariates are time-independent, $\sum_{j=1}^J \sum_{k=1}^{n_j} \int_0^\tau Z_{j,i_k}\, dN^*_{j,k}(t)$ does not depend on the pairing between $N^*_{j,k}(t)$ and Z_{j,i_k}. Since $\sum_{p\in P(n_j)} w_p(j,\beta^m) = 1$, this yields

$$U_{EM}(\beta) = \sum_{j=1}^{J} \sum_{k=1}^{n_j} \int_0^\tau (Z_{j,r_k} - \frac{S_1^{EM}(t,\beta,\beta^m)}{S_0^{EM}(t,\beta,\beta^m)}) dN^*_{j,k}(t), \quad (4)$$

where $(r_1,\ldots,r_{n_j})$ is any ordering of the covariates. This score function for β depends only on the ordering of tie clusters, not on the actual values $T^*_{j,k}$.

Note that $S_v^{EM}(t,\beta,\beta^m)$, $v = 0,1,2$, in the EM-algorithm are calculated for the fixed weights $w_p(j,\beta^m)$ that depend on the iterative values β^m. When converged the score function (4) satisfy that $\hat{\beta} = \beta^m$. Therefore nonparametric maximum likelihood estimator will solve the score equation

$$U(\beta) = \sum_{j=1}^{J} \sum_{k=1}^{n_j} \int_0^\tau (Z_{j,r_k} - \frac{S_1^{EM}(t,\beta,\beta)}{S_0^{EM}(t,\beta,\beta)}) dN^*_{j,k}(t), \quad (5)$$

The asymptotic properties of $\hat{\beta}$ will follow by analyzing the asymptotic properties of this score in the next section.

Based on $\hat{\beta}$ we estimate the cumulative baseline function $\Lambda_0(t)$ by

$$\hat{\Lambda}_0(t) = \int_0^t \frac{1}{S_0^{EM}(s,\hat{\beta},\hat{\beta})}\, dN^*_{\cdot,\cdot}(s)$$

This estimate depends on the actual values $T^*_{j,k}$.

The proposed EM-algorithm is simple to implement. The Newton-Raphson iterative algorithm can be used to solve $U_{EM}(\beta) = 0$ given in (4) to update the estimate of β. The computation is also not intensive, one only needs to consider permutations within each tie. Further, one key property is that the score function $U(\beta)$ has mean 0. To see this we note that one particular ordering of the covariates is the unobserved true $P_j = \{I_1, \ldots, I_{n_j}\}$. Let Z_{j,I_k} be the covariate related to $N^*_{j,k}(t),\ k = 1, \ldots, n_j$. The score (5) can be written as

$$U(\beta) = \sum_{j=1}^{J} \sum_{k=1}^{n_j} \int_0^{\tau} \left(Z_{j,I_k} - \frac{S_1^{EM}(t, \beta, \beta)}{S_0^{EM}(t, \beta, \beta)} \right) dN^*_{j,k}(t).$$

Let F^*_t is the history associated with $(N^*_{j,k}(t), Y^*_{j,k}(t), Z_{j,k}),\ j = 1, .., n, k = 1, .., n_j$, and P_j gives the ordering of covariates. The mean of $U(\beta)$ evaluated at the true value β_0 equals

$$\begin{aligned}
&\sum_{j=1}^{J} \sum_{k=1}^{n_j} E\left[\int_0^{\tau} \left(Z_{j,I_k} - \frac{S_1^{EM}(t, \beta, \beta)}{S_0^{EM}(t, \beta, \beta)} \right) dN^*_{j,k}(t) \right] \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n_j} E\left[E\left(\int_0^{\tau} \left(Z_{j,I_k} - \frac{S_1^{EM}(t, \beta, \beta)}{S_0^{EM}(t, \beta, \beta)} \right) dN^*_{j,k}(t) \middle| F^*_t, P_j \right) \right] \\
&= \sum_{j=1}^{J} \sum_{k=1}^{n_j} E\left[\int_0^{\tau} \left(Z_{j,I_k} - \frac{S_1^{EM}(t, \beta, \beta)}{S_0^{EM}(t, \beta, \beta)} \right) Y^*_{j,k}(t) \exp\left(Z^T_{j,I_k} \beta_0 \right) dt) \right] = 0,
\end{aligned}$$

since

$$\sum_{j=1}^{J} \sum_{k=1}^{n_j} E \left(Y^*_{j,k}(t) \exp\left(Z^T_{j,I_k} \beta_0 \right) \frac{S_1^{EM}(t, \beta, \beta)}{S_0^{EM}(t, \beta, \beta)} \middle| D^* \right) = S_1^{EM}(t, \beta_0, \beta_0),$$

$$\sum_{j=1}^{J} \sum_{k=1}^{n_j} E(Y^*_{j,k}(t) Z_{j,I_k} \exp\left(Z^T_{j,I_k} \beta_0 \right) \Big| D^*) = S_1^{EM}(t, \beta_0, \beta_0).$$

3 Asymptotic properties

The score Eq. 5 is derived under the assumption that each tie cluster consists of either uncensored survival times or censoring times. The EM-algorithm can be extended to deal the situation where censored values are tied with observed survival times, but the procedure becomes more complicated. A simple alternative is to split the tied censoring and observed survival times into two tie clusters and place the tied censoring times right before the tied observed survival times. Each observed failure time contributes one term to the score (5), while a censored failure time cluster contributes only through the at risk sets in $S_v^{EM}(t, \beta, \beta),\ v = 0, 1$. As we noted the score (5) does not depend on the permutations of covariates within each tie cluster. One particular ordering of the covariates is the unobserved true $P_j = \{I_1, \ldots, I_{n_j}\}$. So let Z_{j,I_k} be the covariate related to $N^*_{j,k}(t),\ k = 1, \ldots, n_j$. Thus, the score (5) can be written as

$$U(\beta) = \sum_{j=1}^{J} \sum_{k=1}^{n_j} \int_0^{\tau} \left(Z_{j,I_k} - \frac{S_1^{EM}(t, \beta, \beta)}{S_0^{EM}(t, \beta, \beta)} \right) dN^*_{j,k}(t).$$

Now we denote $\{(N^*_{j,k}(t), Y^*_{j,k}(t), Z_{j,I_k}), j = 1, \ldots, J, k = 1, \ldots, n_j\}$ by its original iid version $\{(N_i(t), Y_i(t), Z_i), i = 1, \ldots, n\}$. Define

$$S_v(t, \beta) = \sum_{j=1}^{J} \sum_{k=1}^{n_j} Z_{j,I_k}^{\otimes v} Y^*_{j,k}(t) \exp(Z_{j,I_k}^T \beta) = \sum_{i=1}^{n} Z_i^{\otimes v} Y_i(t) \exp(Z_i^T \beta)$$

for $v = 0, 1, 2$. Let $s_v(t, \beta) = E(Z_i^{\otimes v} Y_i(t) \exp(Z_i^T \beta))$ and

$$\Sigma(\beta) = \int_0^{\tau} \left(\frac{s_2(t, \beta)}{s_0(t, \beta)} - \left(\frac{s_1(t, \beta)}{s_0(t, \beta)} \right)^{\otimes 2} \right) s_0(t, \beta) \lambda_0(t)\, dt. \tag{6}$$

Also for $v = 0, 1$, let

$$R_v(t, \beta) = \sum_{j=1}^{J} \sum_{p \in P(n_j)} \sum_{k=1}^{n_j} Z_{j,i_k}^{\otimes v} Y^*_{j,k}(t) \exp(Z_{j,i_k}^T \beta) \left(\frac{\partial w_p(j, \beta)}{\partial \beta} \right)^T.$$

Then

$$\begin{aligned} -\,\partial U(\beta)/\partial \beta = & \sum_{j=1}^{J} \sum_{k=1}^{n_j} \int_0^{\tau} \left(\frac{S_2^{EM}(t, \beta, \beta)}{S_0^{EM}(t, \beta, \beta)} - \left(\frac{S_1^{EM}(t, \beta, \beta)}{S_0^{EM}(t, \beta, \beta)} \right)^{\otimes 2} \right) dN^*_{j,k}(t) \\ & + \sum_{j=1}^{J} \sum_{k=1}^{n_j} \int_0^{\tau} \left(\frac{R_1(t, \beta)}{S_0^{EM}(t, \beta, \beta)} - \left(\frac{S_1^{EM}(t, \beta, \beta) R_0(t, \beta)}{(S_0^{EM}(t, \beta, \beta))^2} \right) \right) dN^*_{j,k}(t). \end{aligned} \tag{7}$$

Let $\|z\|$ be the Euclidean metric of a vector or matrix z. The following list of regularity conditions is assumed for the large sample results.

(A.1) $\int_0^{\tau} \lambda_0(t)\, dt < \infty$;
(A.2) There is a constant $b > 2$ and a neighborhood $\mathcal{B}$ of β_0 such that, for $v = 0, 1, 2,\ \sup_{\beta \in \mathcal{B}} E(\|Z_i\|^{vb} \exp(b\|\beta\| \cdot \|Z_i\|)) < \infty$;
(A.3) $P(Y(t) = 1 \text{ for } t \in [0, \tau]) > 0$;
(A.4) $\Sigma(\beta_0)$ is positive definite.

It follows from Theorem 4.1 of Gill and Andersen (1982) that $\sup_{t\in[0,\tau],\beta\in\mathcal{B}} \|n^{-1}S_v(t,\beta)-s_v(t,\beta)\|\xrightarrow{P}0$. The main asymptotic results are presented in the following.

Lemma 1 *Assume that the condition (A.2) is satisfied and that* $\max_{1\leq j\leq J} n_j = O_p(n^a)$ *for* $0\leq a<\frac{1}{2}-\frac{1}{b}$ *holds for the tie sizes. Then*

(*a*) $\sup_{t\in[0,\tau]}|n^{-1/2}(S_v^{EM}(t,\beta,\beta)-S_v(t,\beta))|=o_p(1)$ *for* $\beta\in\mathcal{B}$ *and* $v=0,1,2.$
(*b*) $\sup_{t\in[0,\tau]}|n^{-1}R_v(t,\beta)|=o_p(1)$ *for* $\beta\in\mathcal{B}$ *and* $v=0,1.$

Theorem 1 *Assume that the conditions (A.1)–(A.4) are satisfied and that* $\max_{1\leq j\leq J} n_j = O_p(n^a)$ *for* $0\leq a<\frac{1}{2}-\frac{1}{b}$ *holds for the tie sizes. Then* $\hat\beta\xrightarrow{P}\beta_0$ *and* $\sqrt{n}(\hat\beta-\beta_0)\xrightarrow{\mathcal{D}}N(0,\Sigma^{-1}(\beta_0))$, *as* $n\to\infty$.

Remark 1 As discussed in Sect. 1, the Breslow and Efron approximations lead to the following score functions respectively,

$$U_B(\beta)=\sum_{j=1}^{J}\sum_{k=1}^{n_j}\int_0^\tau (Z_{j,r_k}-\frac{S_1^B(t,\beta,j)}{S_0^B(t,\beta,j)})dN_{j,k}^*(t)$$
$$U_E(\beta)=\sum_{j=1}^{J}\sum_{k=1}^{n_j}\int_0^\tau (Z_{j,r_k}-\frac{S_1^E(t,\beta,j)}{S_0^E(t,\beta,j)})dN_{j,k}^*(t),$$

where $(r_1,\ldots,r_{n_j})$ is any permutation of the indexes $(1,\ldots,n_j)$ for the jth tie cluster,

$$S_v^B(t,\beta,j)=\sum_{j'=j}\sum_{k=1}^{n_{j'}}Z_{j',r_k}^{\otimes v}\exp(Z_{j',r_k}^T\beta)+\sum_{j'\neq j}\sum_{k=1}^{n_{j'}}Z_{j',r_k}^{\otimes v}Y_{j',k}^*(t)\exp(Z_{j',r_k}^T\beta).$$

$$\begin{aligned}S_v^E(t,\beta,j)=&\sum_{j'=j}\sum_{k=1}^{n_{j'}}\frac{n_{j'}-k+1}{n_{j'}}Z_{j',r_k}^{\otimes v}\exp(Z_{j',r_k}^T\beta)\\&+\sum_{j'\neq j}\sum_{k=1}^{n_{j'}}Z_{j',r_k}^{\otimes v}Y_{j',k}^*(t)\exp(Z_{j',r_k}^T\beta).\end{aligned}$$

It is easy to see that the score functions $U_B(\beta)$ and $U_E(\beta)$ do not depend on any particular permutation $(r_1,\ldots,r_{n_j})$. Following the proof of Lemma 1, we also have

$$\max_{1\leq j\leq J}\sup_{t\in[0,\tau]}|n^{-1/2}(S_v^B(t,\beta,j)-S_v(t,\beta))|=o_p(1)$$
$$\max_{1\leq j\leq J}\sup_{t\in[0,\tau]}|n^{-1/2}(S_v^E(t,\beta,j)-S_v(t,\beta))|=o_p(1),$$

for $\beta \in \mathcal{B}$ and $v = 0, 1, 2$. Hence, Theorem 1 also holds for the Breslow and Efron estimators.

4 Simulation studies

In this section, we present some simulation results from an extensive simulation study comparing the proposed EM-algorithm with the existing procedures for dealing with tied survival data under the Cox model. The notable methods for dealing with ties include the methods developed by Breslow and Efron, the Exact method, a simple random break of ties (RB), and the EM-algorithm. All the methods except for the EM-algorithm are implemented using the existing packages from R.

We take the baseline hazard function $\lambda_0(t) = 0.5$ and $\beta = 0, 1, 2$. Four different distributions for the one dimensional covariate x are considered including binary distribution, uniform distribution on [0, 3], normal distribution $N(0, 3^2)$ and exponential distribution with mean equal to 3. These distributions account for some of the practical situations where covariates are binary, approximately uniformly distributed, symmetrical or skewed. Here we choose the binary covariate as .6931×Bernoulli(1, .7), which results in a hazard ratio of 2 in the two sample situation when $\beta = 1$, a scenario considered by Hertz-Picciotto and Rockhill (1997).

We consider simple censorship scenarios: 0% censorship and a light censorship of 30%. The censoring times are generated from an exponential distribution with a parameter selected to give a 30% censoring under each model specification. The ties of the observed failure times are made after the data are generated from the Cox model (possible censored). Specifically, we consider the situations where the maximum size d of the tie in a data set is 2 or 5, which is obtained by grouping neighboring failure/censoring times of the same status up to $d = 2$ or 5 together as a tie after the data is ordered. In the case of no censoring, the number of the failure times in each tie is d except for possibly the last tie. If there is censoring in the data, the number of failure/censoring times in each tie may be less than d. For $d = 5$, for example, if there are four consecutive observed failure times followed by a censored failure time then the size of the tie for observed failure times is 4. If a tie consists of observed failure times then the tied failure time is taken as the median of the group, otherwise the tied censoring time is the first of the group.

The simulations for comparing with different procedures are done for sample size of $n = 100$, $n = 200$ and $n = 300$ with 1000 repetitions. Tables 1–4 list the empirical biases and standard deviations of the estimation for β under each procedure. We also include the results from the full data partial likelihood estimation (Full) in the tables as a benchmark for the effect of information lost due to tie and biases caused by different statistical procedures. Tables 1 and 2 are for tie size $d = 2$ with Table 1 for uncensored data and Table 2 for censored data. Tables 3 and 4 are for tie size $d = 5$ with Table 3 for uncensored data and Table 4 for censored data.

The simulation results from Table 1 to 4 indicate that all procedures perform equally well for the null case $\beta = 0$ and that the EM procedure clearly outperforms the exact

Table 1 Comparison of the bias and standard deviation of estimated β of different procedures for tie size $d = 2$ and 0% censoring

n	β	Full	RB	Breslow	Efron	Exact	EM
Binary covariate							
100	0	−0.0169(0.3256)	−0.0160(0.3252)	−0.0160(0.3167)	−0.0164(0.3230)	−0.0150(0.3315)	−0.0161(0.3276)
	1	0.0044(0.3484)	−0.0014(0.3502)	−0.0258(0.3423)	−0.0055(0.3499)	0.0233(0.3555)	0.0094(0.3535)
	2	0.0338(0.4115)	0.0299(0.4124)	−0.0152(0.4033)	0.0256(0.4111)	0.0701(0.4199)	0.0517(0.4167)
200	0	−0.0083(0.2305)	−0.0084(0.2306)	−0.0082(0.2269)	−0.0085(0.2295)	−0.0078(0.2329)	−0.0084(0.2313)
	1	0.0028(0.2417)	0.0012(0.2425)	−0.0127(0.2398)	−0.0009(0.2426)	0.0134(0.2440)	0.0069(0.2438)
	2	0.0122(0.2763)	0.0109(0.2767)	−0.0115(0.2739)	0.0099(0.2767)	0.0322(0.2790)	0.0226(0.2782)
300	0	−0.0049(0.1793)	−0.0046(0.1787)	−0.0049(0.1771)	−0.0050(0.1785)	−0.0047(0.1803)	−0.0049(0.1794)
	1	0.0021(0.1911)	0.0007(0.1921)	−0.0087(0.1907)	−0.0002(0.1921)	0.0094(0.1930)	0.0051(0.1927)
	2	0.0079(0.2236)	0.0075(0.2238)	−0.0075(0.2223)	0.0071(0.2236)	0.0218(0.2252)	0.0154(0.2244)
Uniform covariate							
100	0	0.0013(0.1237)	0.0013(0.1235)	0.0013(0.1205)	0.0013(0.1229)	0.0014(0.1261)	0.0013(0.1246)
	1	0.0140(0.1526)	0.0122(0.1532)	−0.0075(0.1492)	0.0102(0.1524)	0.0304(0.1554)	0.0222(0.1544)
	2	0.0196(0.2166)	0.0151(0.2168)	−0.0342(0.2077)	0.0101(0.2151)	0.0619(0.2237)	0.0423(0.2209)
200	0	−0.0011(0.0850)	−0.0009(0.0849)	−0.0011(0.0835)	−0.0010(0.0844)	−0.0011(0.0857)	−0.0011(0.0851)
	1	0.0050(0.1049)	0.0043(0.1049)	−0.0057(0.1037)	0.0035(0.1048)	0.0138(0.1059)	0.0095(0.1055)
	2	0.0118(0.1488)	0.0105(0.1490)	−0.0144(0.1459)	0.0091(0.1487)	0.0348(0.1516)	0.0243(0.1505)
300	0	0.0016(0.0673)	0.0015(0.0669)	0.0016(0.0663)	0.0016(0.0668)	0.0016(0.0676)	0.0016(0.0672)
	1	0.0044(0.0865)	0.0041(0.0865)	−0.0027(0.0859)	0.0036(0.0865)	0.0105(0.0871)	0.0076(0.0869)
	2	0.0085(0.1266)	0.0079(0.1265)	−0.0089(0.1249)	0.0070(0.1265)	0.0240(0.1281)	0.0170(0.1275)
Normal covariate							
100	0	−0.0003(0.0346)	−0.0004(0.0345)	−0.0003(0.0336)	−0.0003(0.0343)	−0.0003(0.0352)	−0.0003(0.0348)
	1	0.0090(0.0968)	−0.0041(0.0955)	−0.0515(0.0870)	−0.0179(0.0921)	0.0420(0.1024)	0.0209(0.0990)
	2	0.0173(0.1884)	−0.0891(0.1906)	−0.2163(0.1534)	−0.1222(0.1663)	0.1141(0.2067)	0.0276(0.1892)

Table 1 Continued

n	β	Full	RB	Breslow	Efron	Exact	EM
200	0	0.0007(0.0241)	0.0007(0.0241)	0.0008(0.0237)	0.0008(0.0240)	0.0008(0.0243)	0.0008(0.0242)
	1	0.0046(0.0623)	−0.0007(0.0621)	−0.0263(0.0590)	−0.0063(0.0611)	0.0239(0.0644)	0.0131(0.0633)
	2	0.0088(0.1212)	−0.0651(0.1285)	−0.1107(0.1097)	−0.0486(0.1160)	0.0688(0.1282)	0.0281(0.1237)
300	0	−0.0007(0.0199)	−0.0007(0.0199)	−0.0007(0.0196)	−0.0007(0.0198)	−0.0007(0.0200)	−0.0007(0.0199)
	1	0.0021(0.0513)	−0.0011(0.0515)	−0.0186(0.0498)	−0.0042(0.0510)	0.0159(0.0529)	0.0086(0.0521)
	2	0.0063(0.1004)	−0.0644(0.1093)	−0.0731(0.0944)	−0.0263(0.0984)	0.0514(0.1049)	0.0245(0.1024)
Exponential covariate							
100	0	0.0040(0.0380)	0.0041(0.0379)	0.0041(0.0368)	0.0042(0.0374)	0.0040(0.0385)	0.0042(0.0380)
	1	0.0196(0.1007)	−0.0202(0.1068)	−0.0590(0.0951)	−0.0274(0.0994)	0.0500(0.1061)	0.0263(0.1021)
	2	0.0358(0.1897)	−0.9173(0.3764)	−0.2412(0.1734)	−0.1617(0.1843)	0.1171(0.2040)	0.0119(0.1902)
200	0	0.0021(0.0246)	0.0021(0.0246)	0.0022(0.0242)	0.0022(0.0244)	0.0021(0.0249)	0.0022(0.0247)
	1	0.0068(0.0652)	−0.0321(0.0759)	−0.0348(0.0619)	−0.0155(0.0636)	0.0251(0.0675)	0.0130(0.0655)
	2	0.0119(0.1246)	−1.070(0.3122)	−0.1362(0.1166)	−0.0839(0.1212)	0.0627(0.1305)	0.0140(0.1228)
300	0	0.0017(0.0198)	0.0017(0.0198)	0.0017(0.0196)	0.0017(0.0198)	0.0016(0.0200)	0.0017(0.0199)
	1	0.0045(0.0538)	−0.0395(0.0668)	−0.0239(0.0523)	−0.0095(0.0534)	0.0179(0.0553)	0.0099(0.0543)
	2	0.0073(0.1035)	−1.1604(0.2645)	−0.0937(0.1004)	−0.0536(0.1035)	0.0452(0.1073)	0.0179(0.1042)

Table 2 Comparison of the bias and standard deviation of estimated β of different procedures for tie size $d = 2$ and 30% censoring

n	β	Full	RB	Breslow	Efron	Exact	EM
Binary covariate							
100	0	−0.0086(0.3881)	−0.0072(0.3885)	−0.0070(0.3807)	−0.0071(0.3868)	−0.0059(0.3950)	−0.0069(0.3913)
	1	0.0085(0.4144)	0.0054(0.4166)	−0.0146(0.4055)	0.0025(0.4136)	0.0282(0.4231)	0.0155(0.4192)
	2	0.0372(0.5041)	0.0324(0.5044)	−0.0089(0.4895)	0.0251(0.4987)	0.0720(0.5103)	0.0506(0.5068)
200	0	0.0037(0.2736)	0.0043(0.2739)	0.0039(0.2709)	0.0040(0.2737)	0.0044(0.2765)	0.0041(0.2754)
	1	0.0136(0.2808)	0.0129(0.2807)	0.0005(0.2772)	0.0106(0.2803)	0.0246(0.2838)	0.0176(0.2822)
	2	0.0275(0.3336)	0.0256(0.3348)	0.0045(0.3309)	0.0239(0.3345)	0.0471(0.3381)	0.0360(0.3363)
300	0	−0.0011(0.2163)	−0.0009(0.2162)	−0.0013(0.2143)	−0.0012(0.2157)	−0.0011(0.2175)	−0.0012(0.2167)
	1	0.0060(0.2275)	0.0046(0.2271)	−0.0042(0.2255)	0.0035(0.2272)	0.0127(0.2291)	0.0085(0.2282)
	2	0.0093(0.2637)	0.0087(0.2637)	−0.0066(0.2623)	0.0070(0.2639)	0.0225(0.2656)	0.0154(0.2650)
Uniform covariate							
100	0	−0.0012(0.1449)	−0.0018(0.1451)	−0.0017(0.1421)	−0.0018(0.1446)	−0.0017(0.1473)	−0.0018(0.1462)
	1	0.0198(0.1822)	0.0187(0.1812)	0.0022(0.1790)	0.0169(0.1819)	0.0344(0.1855)	0.0272(0.1839)
	2	0.0408(0.2477)	0.0383(0.2474)	−0.0009(0.2397)	0.0351(0.2463)	0.0783(0.2545)	0.0608(0.2515)
200	0	−0.0035(0.0988)	−0.0037(0.0988)	−0.0036(0.0979)	−0.0036(0.0988)	−0.0036(0.0999)	−0.0036(0.0994)
	1	0.0050(0.1187)	0.0047(0.1186)	−0.0036(0.1175)	0.0042(0.1185)	0.0131(0.1197)	0.0092(0.1192)
	2	0.0099(0.1654)	0.0091(0.1654)	−0.0107(0.1627)	0.0079(0.1651)	0.0289(0.1678)	0.0196(0.1666)
300	0	0.0019(0.0822)	0.0018(0.0822)	0.0018(0.0815)	0.0018(0.0821)	0.0019(0.0827)	0.0019(0.0824)
	1	0.0049(0.0979)	0.0047(0.0980)	−0.0010(0.0974)	0.0044(0.0980)	0.0104(0.0985)	0.0077(0.0983)
	2	0.0107(0.1353)	0.0104(0.1353)	−0.0031(0.1338)	0.0098(0.1351)	0.0237(0.1367)	0.0177(0.1360)
Normal covariate							
100	0	−0.0001(0.0419)	−0.0001(0.0419)	−0.0001(0.0411)	−0.0001(0.0418)	−0.0001(0.0426)	−0.0001(0.0422)
	1	0.0086(0.1092)	0.0019(0.1078)	−0.0335(0.1010)	−0.0057(0.1057)	0.0369(0.1148)	0.0242(0.1120)
	2	0.0156(0.2087)	−0.0760(0.2045)	−0.1647(0.1770)	−0.0830(0.1898)	0.1015(0.2263)	0.0424(0.2121)

Table 2 Continued

n	β	Full	RB	Breslow	Efron	Exact	EM
200	0	0.0003(0.0288)	0.0004(0.0288)	0.0004(0.0284)	0.0003(0.0287)	0.0004(0.0291)	0.0003(0.0289)
	1	0.0053(0.0706)	0.0029(0.0706)	−0.0161(0.0677)	0.0002(0.0697)	0.0211(0.0729)	0.0148(0.0721)
	2	0.0126(0.1372)	−0.0653(0.1455)	−0.0777(0.1266)	−0.0251(0.1326)	0.0645(0.1448)	0.0367(0.1401)
300	0	0.0000(0.0234)	0.0000(0.0234)	0.0000(0.0232)	0.0000(0.0233)	0.0000(0.0235)	0.0000(0.0234)
	1	0.0026(0.0599)	0.0013(0.0598)	−0.0115(0.0582)	−0.0001(0.0594)	0.0138(0.0613)	0.0093(0.0606)
	2	0.0055(0.1148)	−0.0771(0.1289)	−0.0549(0.1090)	−0.0164(0.1128)	0.0423(0.1194)	0.0239(0.1171)
Exponential covariate							
100	0	0.0030(0.0454)	0.0029(0.0453)	0.0030(0.0442)	0.0031(0.0449)	0.0029(0.0459)	0.0031(0.0455)
	1	0.0145(0.1083)	−0.0330(0.1210)	−0.0687(0.1008)	−0.0389(0.1049)	0.0437(0.1145)	0.0189(0.1095)
	2	0.0329(0.2151)	−0.9675(0.3832)	−0.2671(0.1921)	−0.1931(0.2028)	0.1127(0.2295)	−0.0093(0.2101)
200	0	0.0009(0.0299)	0.0009(0.0299)	0.0009(0.0295)	0.0009(0.0298)	0.0008(0.0302)	0.0008(0.0300)
	1	0.0038(0.0703)	−0.0412(0.0847)	−0.0405(0.0669)	−0.0217(0.0687)	0.0221(0.0723)	0.0094(0.0705)
	2	0.0063(0.1361)	−1.1413(0.3042)	−0.1554(0.1292)	−0.1046(0.1341)	0.0568(0.1422)	0.0039(0.1358)
300	0	0.0012(0.0241)	0.0013(0.0241)	0.0012(0.0239)	0.0013(0.0240)	0.0012(0.0243)	0.0012(0.0242)
	1	0.0049(0.0608)	−0.0435(0.0731)	−0.0253(0.0583)	−0.0110(0.0596)	0.0184(0.0619)	0.0097(0.0610)
	2	0.0130(0.1166)	−1.2131(0.2628)	−0.0986(0.1100)	−0.0582(0.1137)	0.0520(0.1201)	0.0220(0.1162)

Table 3 Comparison of the bias and standard deviation of estimated β of different procedures for tie size $d = 5$ and 0% censoring

n	β	Full	RB	Breslow	Efron	Exact	EM
Binary covariate							
100	0	−0.0169(0.3256)	−0.0147(0.3264)	−0.0147(0.2981)	−0.0163(0.3176)	−0.0129(0.3492)	−0.0153(0.3368)
	1	0.0044(0.3484)	−0.0152(0.3527)	−0.1047(0.3220)	−0.0431(0.3485)	0.0682(0.3766)	0.0229(0.3668)
	2	0.0338(0.4115)	0.0101(0.4167)	−0.1551(0.3855)	−0.0145(0.4170)	0.1616(0.4474)	0.0972(0.4365)
200	0	−0.0083(0.2305)	−0.0072(0.2318)	−0.0072(0.2190)	−0.0081(0.2274)	–	−0.0076(0.2347)
	1	0.0028(0.2417)	−0.0068(0.2463)	−0.0564(0.2338)	−0.0178(0.2459)	–	0.0161(0.2507)
	2	0.0122(0.2763)	0.0077(0.2777)	−0.0797(0.2671)	0.0006(0.2781)	–	0.0530(0.2840)
300	0	−0.0049(0.1793)	−0.0042(0.1782)	−0.0042(0.1720)	−0.0046(0.1763)	–	−0.0044(0.1804)
	1	0.0021(0.1911)	−0.0019(0.1934)	−0.0380(0.1875)	−0.0086(0.1936)	–	0.0138(0.1959)
	2	0.0090(0.2236)	0.0055(0.2246)	−0.0533(0.2188)	0.0027(0.2243)	–	0.0366(0.2275)
Uniform covariate							
100	0	0.0013(0.1237)	0.0003(0.1239)	0.0008(0.1134)	0.0009(0.1208)	0.0012(0.1327)	0.0010(0.1281)
	1	0.0140(0.1526)	0.0045(0.1527)	−0.0675(0.1399)	−0.0058(0.1510)	0.0733(0.1638)	0.0453(0.1594)
	2	0.0196(0.2166)	−0.0046(0.2166)	−0.1831(0.1859)	−0.0310(0.2100)	0.1714(0.2461)	0.1048(0.2351)
200	0	−0.0011(0.0850)	−0.0013(0.0851)	−0.0012(0.0806)	−0.0011(0.0836)	–	−0.0011(0.0863)
	1	0.0050(0.1049)	0.0024(0.1050)	−0.0360(0.1000)	−0.0017(0.1044)	–	0.0230(0.1072)
	2	0.0118(0.1488)	0.0043(0.1482)	−0.0901(0.1373)	−0.0036(0.1473)	–	0.0597(0.1549)
300	0	0.0016(0.0673)	0.0011(0.0674)	0.0010(0.0650)	0.0011(0.0667)	–	0.0012(0.0682)
	1	0.0044(0.0866)	0.0026(0.0867)	−0.0234(0.0842)	0.0005(0.0866)	–	0.0167(0.0880)
	2	0.0085(0.1266)	0.0047(0.1264)	−0.0596(0.1199)	0.0007(0.1258)	–	0.0416(0.1299)
Normal covariate							
100	0	−0.0003(0.0346)	−0.0001(0.0348)	−0.0002(0.0318)	−0.0002(0.0338)	−0.0002(0.0373)	−0.0002(0.0359)
	1	0.0090(0.0968)	−0.0446(0.0974)	−0.1937(0.0735)	−0.0983(0.0862)	0.1178(0.1192)	0.0556(0.1076)
	2	0.0173(0.1884)	−0.2611(0.1903)	−0.6795(0.1230)	−0.4745(0.1490)	0.3216(0.2511)	0.0632(0.1952)

Table 3 continued

n	β	Full	RB	Breslow	Efron	Exact	EM
200	0	0.0007(0.0241)	0.0007(0.0243)	0.0007(0.0230)	0.0007(0.0239)	–	0.0007(0.0246)
	1	0.0046(0.0623)	−0.0171(0.0615)	−0.1056(0.0531)	−0.0400(0.0593)	–	0.0389(0.0665)
	2	0.0088(0.1212)	−0.1431(0.1376)	−0.3930(0.0956)	−0.2244(0.1118)	–	0.0869(0.1313)
300	0	−0.0007(0.0199)	−0.0008(0.0199)	−0.0007(0.0191)	−0.0007(0.0196)	–	−0.0007(0.0201)
	1	0.0021(0.0513)	−0.0104(0.0512)	−0.0743(0.0464)	−0.0245(0.0504)	–	0.0280(0.0543)
	2	0.0063(0.1004)	−0.1374(0.1211)	−0.2805(0.0842)	−0.1377(0.0959)	–	0.0760(0.1066)
Exponential covariate							
100	0	0.0040(0.0380)	0.0045(0.0382)	0.0041(0.0349)	0.0045(0.0370)	0.0040(0.0408)	0.0044(0.0393)
	1	0.0196(0.1007)	−0.0809(0.1213)	−0.2724(0.1183)	−0.1895(0.1314)	0.1170(0.1211)	0.0519(0.1084)
	2	0.0358(0.1897)	−1.1216(0.3460)	−0.8317(0.2236)	−0.6667(0.2523)	0.2898(0.2491)	0.0166(0.2016)
200	0	0.0021(0.0246)	0.0020(0.0247)	0.0020(0.0235)	0.0021(0.0243)	–	0.0020(0.0251)
	1	0.0068(0.0652)	−0.0590(0.0860)	−0.1570(0.0716)	−0.0979(0.0773)	–	0.0335(0.0685)
	2	0.0119(0.1246)	−1.30674(0.2334)	−0.5161(0.1529)	−0.3784(0.1670)	–	0.0421(0.1284)
300	0	0.0017(0.0198)	0.0018(0.0198)	0.0017(0.0192)	0.0018(0.0196)	–	0.0017(0.0201)
	1	0.0045(0.0538)	−0.0640(0.0787)	−0.1093(0.0579)	−0.0628(0.0616)	–	0.0262(0.0560)
	2	0.0073(0.1034)	−1.3905(0.1880)	−0.3728(0.1245)	−0.2568(0.1344)	–	0.0488(0.1088)

Table 4 Comparison of the bias and standard deviation of estimated β of different procedures for tie size $d = 5$ and 30% censoring

n	β	Full	RB	Breslow	Efron	Exact	EM
Binary covariate							
100	0	−0.0086(0.3881)	−0.0066(0.3877)	−0.0066(0.3656)	−0.0082(0.3819)	−0.0049(0.4058)	−0.0078(0.3956)
	1	0.0085(0.4144)	0.0032(0.4191)	−0.0584(0.3863)	−0.0113(0.4074)	0.0643(0.4375)	0.0303(0.4251)
	2	0.0372(0.5041)	0.0258(0.5029)	−0.0837(0.4765)	0.0038(0.4979)	0.1306(0.5266)	0.0756(0.5172)
200	0	0.0037(0.2736)	0.0038(0.2743)	0.0037(0.2646)	0.0038(0.2719)	–	0.0040(0.2772)
	1	0.0136(0.2808)	0.0102(0.2828)	−0.0251(0.2702)	0.0041(0.2793)	–	0.0266(0.2855)
	2	0.0275(0.3336)	0.0245(0.3366)	−0.0363(0.3254)	0.0167(0.3346)	–	0.0522(0.3399)
300	0	−0.0011(0.2163)	−0.0009(0.2163)	−0.0012(0.2109)	−0.0013(0.2145)	–	−0.0012(0.2174)
	1	0.0060(0.2275)	0.0025(0.2272)	−0.0232(0.2220)	−0.0008(0.2268)	–	0.0148(0.2303)
	2	0.0093(0.2637)	0.0073(0.2647)	−0.0347(0.2601)	0.0029(0.2641)	–	0.0269(0.2661)
Uniform covariate							
100	0	−0.0012(0.1449)	−0.0019(0.1438)	−0.0013(0.1363)	−0.0014(0.1426)	−0.0013(0.1510)	−0.0014(0.1477)
	1	0.0198(0.1823)	0.0166(0.1827)	−0.0309(0.1733)	0.0115(0.1812)	0.0621(0.1912)	0.0432(0.1876)
	2	0.0408(0.2477)	0.0336(0.2473)	−0.0839(0.2252)	0.0229(0.2427)	0.1526(0.2685)	0.1038(0.2591)
200	0	−0.0035(0.0988)	−0.0033(0.0993)	−0.0032(0.0959)	−0.0033(0.0984)	–	−0.0033(0.1003)
	1	0.0050(0.1187)	0.0044(0.1189)	−0.0202(0.1156)	0.0028(0.1185)	–	0.0180(0.1205)
	2	0.0099(0.1654)	0.0081(0.1649)	−0.0522(0.1578)	0.0050(0.1645)	–	0.0419(0.1692)
300	0	0.0019(0.0822)	0.0017(0.0823)	0.0017(0.0802)	0.0017(0.0818)	–	0.0017(0.0829)
	1	0.0049(0.0979)	0.0045(0.0981)	−0.0124(0.0965)	0.0035(0.0980)	–	0.0136(0.0990)
	2	0.0107(0.1353)	0.0099(0.1349)	−0.0320(0.1309)	0.0079(0.1346)	–	0.0324(0.1372)
Normal covariate							
100	0	−0.0001(0.0419)	−0.0002(0.0418)	−0.0003(0.0394)	−0.0004(0.0411)	−0.0004(0.0437)	−0.0004(0.0426)
	1	0.0086(0.1092)	−0.0216(0.1079)	−0.1328(0.0901)	−0.0531(0.1011)	0.0925(0.1271)	0.0592(0.1189)
	2	0.0156(0.2087)	−0.1954(0.2166)	−0.5531(0.1599)	−0.3668(0.1882)	0.2720(0.2732)	0.1029(0.2272)

Table 4 cotinued

n	β	Full	RB	Breslow	Efron	Exact	EM
200	0	0.0003(0.0288)	0.0004(0.0288)	0.0004(0.0278)	0.0004(0.0286)	–	0.0004(0.0291)
	1	0.0053(0.0706)	−0.0059(0.0697)	−0.0698(0.0630)	−0.0185(0.0682)	–	0.0365(0.0745)
	2	0.0126(0.1372)	−0.1272(0.1596)	−0.3086(0.1158)	−0.1610(0.1322)	–	0.0924(0.1473)
300	0	0.0000(0.0234)	0.0000(0.0233)	0.0000(0.0228)	0.0000(0.0233)	–	0.0000(0.0236)
	1	0.0026(0.0599)	−0.0039(0.0598)	−0.0483(0.0556)	−0.0107(0.0589)	–	0.0250(0.0627)
	2	0.0055(0.1148)	−0.1381(0.1480)	−0.2172(0.1021)	−0.0999(0.1135)	–	0.0688(0.1230)
Exponential covariate							
100	0	0.0030(0.0454)	0.0031(0.0454)	0.0030(0.0426)	0.0032(0.0446)	0.0028(0.0473)	0.0032(0.0463)
	1	0.0145(0.1083)	−0.0909(0.1375)	−0.2696(0.1301)	−0.1996(0.1416)	0.0973(0.1249)	0.0354(0.1108)
	2	0.0323(0.2151)	−1.1586(0.3571)	−0.8452(0.2419)	−0.7055(0.2668)	0.2445(0.2642)	−0.0320(0.2066)
200	0	0.0009(0.0299)	0.0009(0.0299)	0.0010(0.0290)	0.0010(0.0297)	–	0.0009(0.0303)
	1	0.0038(0.0703)	−0.0707(0.1008)	−0.1632(0.0817)	−0.1114(0.0870)	–	0.0247(0.0724)
	2	0.0063(0.1361)	−1.355(0.2225)	−0.5442(0.1727)	−0.4232(0.1864)	–	0.0162(0.1392)
300	0	0.0012(0.0241)	0.0012(0.0241)	0.0012(0.0235)	0.0012(0.0240)	–	0.0012(0.0243)
	1	0.0049(0.0608)	−0.0728(0.0927)	−0.1116(0.0654)	−0.0697(0.0691)	–	0.0231(0.0628)
	2	0.0130(0.1166)	−1.4236(0.1797)	−0.3905(0.1378)	−0.2847(0.1476)	–	0.0428(0.1201)

procedure. The results for the exact procedure are eliminated from the tables for tie size $d = 5$ and n greater than 200 since the procedure is extremely time consuming. The procedure using a random break of tie seems perform well for binary and uniform covariates. However, the bias increases as β increases. Our simulation (not reported here) shows that this procedure breaks down for large values of β, say $\beta = 6$, resulting in very large biases. It also performs very poorly for the exponential covariate distributions. The Breslow procedure also works well with binary and uniform covariate distributions. But it has larger biases than the EM for normal and exponential covariate distributions, especially for larger ties, similar to RB procedure. Our simulation study shows that Efron procedure has the best performance of the existing methods dealing tied survival data, which is consistent with the findings of Hertz-Picciotto and Rockhill (1997). Compared with the EM procedure, Efron procedure works better with binary and uniform covariate distributions. It has generally larger biases than the EM for normal and exponential covariate distributions, particularly for large tie and small sample size. This effect becomes more evident when the coefficient β increases. The simulation results in Table 1–4 suggest that the proposed EM-algorithm is a reliable procedure to use in practice because of its overall well performance across different covariate distributions, β values, sample sizes, tie sizes and censorship status.

5 Summary

We have studied several key approaches for dealing with ties in survival data using Cox regression model. Our simulation clearly show that the Efron procedure is the best choice among the implemented procedures. The proposed new approach using EM-algorithm based on the likelihood leads to a score function that is unbiased. The performances of the EM-based approach in the simulations are good and robust in all situations. Our method is comparable to the Efron procedure in most situations and it does better for normal and exponential covariate distributions for smaller sample size with larger tie size. This phenonmenon becomes more evident when the coefficient β increases. The overall well performance of the EM-based procedure across different covariate distributions, β values, sample sizes, tie sizes and censorship status make it a reliable procedure to use in practice.

In addition to the numerical study we also showed that all available methods have the same asymptotic properties and that all methods are asymptotically fully efficient. Quite surprisingly the tied survival data result in no loss in asymptotic efficiency when compared to the un-tied survival data.

Appendix

Proof of Lemma 1 Note that at each time t, if the tied cluster j failed before or after t, then $Y^*_{j,k}(t) = 0$ or 1 for all k depending on whether the jth cluster fails before t or after t, in which case $\sum_k Z^{\otimes v}_{j,i_k} Y^*_{j,k}(t) \exp(Z^T_{j,i_k}\beta)$ does not depend on the permutation of indices. And since $\sum_{p\in P(n_j)} w_p(j,\beta^m) = 1$ and $\sum_{p\in P(n_j)} \frac{\partial w_p(j,\beta^m)}{\partial\beta} = 0$, it follows that

$$\sum_{p\in P(n_j)} w_p(j,\beta^m) \sum_{k=1}^{n_j} Z_{j,i_k}^{\otimes v} Y_{j,k}^*(t)\exp(Z_{j,i_k}^T\beta) = \sum_{k=1}^{n_j} Z_{j,I_k}^{\otimes v} Y_{j,k}^*(t)\exp(Z_{j,I_k}^T\beta) \quad (8)$$

$$\sum_{p\in P(n_j)} \sum_{k=1}^{n_j} Z_{j,i_k}^{\otimes v} Y_{j,k}^*(t)\exp(Z_{j,i_k}^T\beta)\left(\frac{\partial w_p(j,\beta^m)}{\partial\beta}\right)^T = 0. \quad (9)$$

The Eqs. 8 and 9 do not hold for at most one tie cluster at each time under the assumption $\max_k(T_{j,k}^*) < \min_k(T_{l,k}^*)$ for any $j < l$. Hence the jth term in $S_v^{EM}(t,\beta,\beta)$ and the jth term in $S_v(t,\beta)$ are equal except for at most one cluster.

The assertion (a) follows by proving that both the terms, $\sum_{p\in P(n_j)} w_p(j,\beta^m)$ $\sum_{k=1}^{n_j} Z_{j,i_k}^{\otimes v} Y_{j,k}^*(t)\ \exp(Z_{j,i_k}^T\beta)$ and $\sum_{k=1}^{n_j} Z_{j,r_k}^{\otimes v} Y_{j,k}^*(t)\exp(Z_{j,r_k}^T\beta)$, are at the order of $O_p(n^{1/2-\delta})$, for some $\delta > 0$, uniformly in $j\in\{1,\ldots,J\}$, $p\in P(n_j)$ and $t\in[0,\tau]$, for $v = 0, 1, 2$. Note that both of the terms are bounded by $(\max_j n_j)\|Z\|_{(n)}^v \exp(\|Z\|_{(n)}\|\beta\|)$, where $\|Z\|_{(n)}$ is the maximum of $\|Z_i\|$, $i = 1,\ldots,n$. Since $\max_j n_j = O_p(n^a)$, it suffices to show that $n^a\|Z\|_{(n)}^v \exp(\|Z\|_{(n)}\|\beta\|) = O_p(n^{1/2-\delta})$, for some $\delta > 0$. Let $F(x)$ be the distribution function of $\|Z_i\|^v \exp(\|Z_i\|\|\beta\|)$. Then

$$P(\|Z\|_{(n)}^v \exp(\|Z\|_{(n)}\|\beta\| > n^{1/2-\delta-a}) = P(\max_{1\le i\le n} \|Z_i\|^v \exp(\|Z_i\|\|\beta\| > n^{1/2-\delta-a})$$
$$= 1 - \left[1 - \frac{n(1-F(n^{1/2-\delta-a}))}{n}\right]^n. \quad (10)$$

Since $n(1-F(n^{1/2-\delta-a})) \le n^{1+b(-1/2+\delta+a)} \int_{n^{1/2-\delta-a}}^{\infty} x^b dF(x) = o(n^{1+b(-1/2+\delta+a)})$ $= o(1)$, by choosing $0 < \delta < \frac{1}{2} - \frac{1}{b} - a$, the right hand side of (10) goes to 0 as $n\to\infty$. Hence $\|Z\|_{(n)}^v \exp(\|Z\|_{(n)}\|\beta\|) = O_p(n^{1/2-\delta-a})$, for $v = 0, 1, 2$. This proves part (a).

To prove part (b), following the argument that leads to (9), we note that

$$\sup_{0\le t\le\tau} \|R_v(t,\beta)\| \le \sup_{0\le t\le\tau} \max_{1\le j\le J} \left\| \sum_{p\in P(n_j)} \sum_{k=1}^{n_j} Z_{j,i_k}^{\otimes v} Y_{j,k}^*(t)\exp(Z_{j,i_k}^T\beta)\left(\frac{\partial w_p(j,\beta)}{\partial\beta}\right)^T \right\|$$
$$\le (\max_{1\le j\le J} n_j)\|Z\|_{(n)}^v \exp(\|Z\|_{(n)}\|\beta\|) \max_{1\le j\le J} \sum_{p\in P(n_j)} \left\|\frac{\partial w_p(j,\beta)}{\partial\beta}\right\|$$
$$= O_p(n^{1/2-\delta}) \max_{1\le j\le J} \sum_{p\in P(n_j)} \left\|\frac{\partial w_p(j,\beta)}{\partial\beta}\right\|.$$

Since $w_p(j, \beta^m)$ is the Cox partial likelihood for the jth cluster of a given permutation,

$$\frac{\partial w_p(j,\beta)}{\partial\beta} = \frac{\partial \exp(\log w_p(j,\beta))}{\partial\beta} = w_p(j,\beta)\frac{\partial \log w_p(j,\beta)}{\partial\beta}$$
$$= w_p(j,\beta)\sum_{k=1}^{n_j}\int_0^\tau\left(Z_{j,i_k} - \frac{\sum_{k=1}^{n_j} Z_{j,i_k}Y^*_{j,k}(t)\exp(Z^T_{j,i_k}\beta)}{\sum_{k=1}^{n_j} Y^*_{j,k}(t)\exp(Z^T_{j,i_k}\beta)}\right)dN^*_{j,k}(t).$$

We have

$$\max_{1\le j\le J}\sum_{p\in P(n_j)}\left\|\frac{\partial w_p(j,\beta)}{\partial\beta}\right\| \le 2(\max_{1\le j\le J} n_j)\|Z\|_{(n)}\sum_{p\in P(n_j)} w_p(j,\beta) = O_p(n^{1/2-\delta}).$$

This is followed by $\sup_{0\le t\le\tau}\|R_v(t,\beta)\| = O_p(n^{1-2\delta})$. This completes the proof. □

Proof of Theorem 1 By Lemma 1, $n^{-1}U(\beta) = n^{-1/2}\sum_{i=1}^n\int_0^\tau(Z_i - \frac{S_1(t,\beta)}{S_0(t,\beta)})\, dN_i(t) + o_p(1)$, for $\beta\in\mathcal{B}$. Under the conditions (A.1)–(A.4) and applying the strong law of large numbers, we have $n^{-1}U(\beta)\xrightarrow{P}u(\beta)$, where $u(\beta) = \int_0^\tau(s_1(t,\beta_0) - s_0(t,\beta_0)s_1(t,\beta)/s_0(t,\beta))\lambda_0(t)\,dt$, for $\beta\in\mathcal{B}$. Since $\Sigma(\beta_0)$ is positive definite, β_0 is the unique zero of $u(\beta)$. It follows by Lemma 5.10 of van der Vaart (1998) that $\hat\beta\xrightarrow{P}\beta_0$.

Now we prove the asymptotic normality. Since $U(\hat\beta) = 0$, we have $\sqrt{n}(\hat\beta - \beta_0) = (-n^{-1}\partial U(\beta^*)/\partial\beta)^{-1}\, n^{-1/2}U(\beta_0)$, where β^* is on the line segment between $\hat\beta$ and β_0. It follows by Lemma 1 that $n^{-1/2}U(\beta_0) = n^{-1/2}\sum_{i=1}^n\int_0^\tau(Z_i - \frac{S_1(t,\beta_0)}{S_0(t,\beta_0)})\, dN_i(t) + o_p(1)$, and $-n^{-1}\partial U(\beta)/\partial\beta \xrightarrow{P}\Sigma(\beta)$ in a neighborhood of β_0 by the strong law of large numbers. A routine application of martingale central limit theorem shows $n^{-1/2}\sum_{j=1}^n\int_0^\tau(Z_i - \frac{S_1(t,\beta_0)}{S_0(t,\beta_0)})\, dN_i(t)\xrightarrow{\mathcal{D}} N(0,\Sigma(\beta_0))$. By Slusky's theorem, we have $\sqrt{n}(\hat\beta - \beta_0)\xrightarrow{\mathcal{D}}N(0,\Sigma^{-1}(\beta_0))$, as $n\to\infty$. □

Acknowledgements We thank the two referees for their valuable comments. The research work of Yanqing Sun was partially supported by NSF grants DMS-0304922 and DMS-0604576 and NIH grant 2 RO1 AI054165-04. Part of the work was done while Thomas Scheike visited the Centre for Advanced Study in Oslo.

References

Andersen PK, Borgan Ø, Gill RD, Keiding N (1993) Statistical Models Based on Counting Processes. Springer-Verlag, New York

Breslow NE (1972) Discussion on the paper by D.R. Cox. J Roy Statist Soc Ser B 34:216–217

Cox DR (1972) Regression models and life tables (with discussion). J Roy Statist Soc B 34:187–220

Efron B (1977) The efficiency of Cox's likelihood function for censored data. J Amer Statist Assoc 72:557–565

Hertz-Picciotto I, Rockhill B (1997) Validity and efficiency of approximation methods for tied survival times in Cox regression. Biometrics 53:1151–1156

Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data. Wiley, New York

Peto R (1972) Contribution to the discussion of the paper by D. R. Cox. J Roy Statist Soc Ser B 34:205–207

Satten GA (1996) Rank-based inference in the proportional hazards model for interval censored data. Biometrika 83:355–370

Therneau TM, Grambsch PM (2000) Modeling survival data: extending the Cox model. Springer-Verlag, New York