EM algorithm for analyzing right-censored survival data under the semiparametric proportional odds model

Lu Wang* and Lianming Wang

Department of Statistics, University of South Carolina, 1523 Greene Street, Columbia, South Carolina, 29208, U.S.A.

**Email: lw3@email.sc.edu*

Abstract: The semiparametric proportional odds (PO) model is a popular alternative to Cox's proportional hazards model for analyzing survival data. Although many approaches have been proposed for this topic in the literature, most of the existing approaches have been found computationally expensive and difficult to implement. In this article, we proposed a novel and easy-to-implement approach based on an expectation-maximization (EM) algorithm for analyzing right-censored data. Specifically, our approach adopts monotone splines of Ramsay (1988) to approximate the cumulative baseline odds function, which allows us to reduce the number of unknown parameters from infinite to finite. Upon the use of monotone splines, a novel EM algorithm is developed to estimate the regression coefficients and spline coefficients simultaneously. The EM algorithm is simple to implement, robust to initial values, and converges fast as it only involves solving a low-dimensional estimating equation for the regression parameters and then updating the spline coefficients in simple closed form at each iteration. The regression parameter estimate is asymptotically normally distributed and its variance estimate is obtained in closed form. Simulation studies suggest that our proposed method has excellent performance in estimating both regression parameters and cumulative baseline odds and survival functions, even when the right censoring rate is very high. Our method is applied to a large dataset about prostate cancer screening from the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial conducted by the United States National Cancer Institute.

Keywords: Semiparametric regression; EM algorithm; Monotone splines; Proportional odds model; Survival analysis.

1 Introduction

The semiparametric proportional odds (PO) model is a popular alternative to Cox's proportional hazards (PH) model for analyzing survival data due to its nice model properties and good interpretation of regression coefficients in terms of log odds ratios. The PO model was first introduced by McCullagh (1977) and McCullagh (1980) for analyzing ordinal response data. Later Bennett (1983a; 1983b) and Pettitt (1984) generalized the PO model to the field of survival analysis. Unlike the PH model, in a two-sample problem, the PO model implies that the ratio of the hazards converges to unity as time goes to infinity. For this reason, Bennett (1983a) suggested to use the PO model to demonstrate an effective cure, and Murphy et al. (1997) suggesed to use the PO model when the morbidity rates converge with time. Existing work under the PO model for right-censored data include Banerjee & Dey (2005), Hanson & Yang (2007), and Banerjee et al. (2007) from Bayesian perspectives and Bennett (1983a), Murphy et al. (1997), Shen (1998), and Royston & Parmar (2002) from frequentist's prespectives. The PO model has also been popularly adopted to analyze interval-censored survival data in the literature, and existing approaches in this category include Rossini & Tsiatis (1996), Huang & Rossini (1997), Robinowitz et al. (2000), Wang & Dunson (2011), and Lin & Wang (2011) among others.

In spite of the popularity of the PO model, our investigation has found that the existing approaches for right-censored data under the PO model are computationally expensive and difficulty to implement. In general, earlier work in the semiparametric framework mainly focused on estimating regression coefficients and treating baseline odds function as an infinite dimensional nuisance parameter, and Murphy et al. (1997) was the first to allow joint estimation of the regression parameters and the baseline odds function. Murphy et al. (1997) approximated the baseline odds function with a step function that has jumps at all of the exactly observed failure times, and for this reason, their method became less effective and problematic when the failure times are observed continuously. Shen (1998) solved this problem by adopting monotone splines with variable orders and knots to approximate the baseline odds function and proposed a sieve maximum likelihood estimate for right-censored and case 2 interval-censored data under the PO model. However, Shen (1998)'s method was very complicated and hard to implement, as commented in Royston & Parmar (2002). Royston & Parmar (2002) instead proposed to use natural cubic splines to approximate the log baseline cumulative odds function. However, the use of natural cubic spline does not guarantee the global monotonicity of the baseline cumulative odds function. In addition, as mentioned in their paper that the estimation is sensitive to the initial values.

In this article, we propose a novel estimation method for analyzing right-censored data under the PO model via an expectation-maximization (EM) algorithm. Specifically we approximate the baseline cumulative odds function with monotone splines of Ramsay (1988), and consequently the baseline odds function is approximated by the corresponding M splines with the same set of spline coefficients. A novel data augmentation is proposed for the derivation of our EM algorithm that allows to estimate the regression parameters and spline coefficients jointly. The resulting EM algorithm is straightforward to implement as it only involves solving a low-dimensional estimating equation for the regression parameters and then updating all the spline coefficients in simple closed form at each iteration. Our EM algorithm is found to be robust to initial values and fast to converge, and the variance of the regression parameter estimator is also estimated in closed form. In addition, our method is found to work well even when the right-censoring rate is very high, in which cases direct optimization method often fails.

The remainder of this paper is organized as follows. Section 2 introduces the notations, model, observed data and likelihood. Section 3 proposes to use the splines to approximate the baseline (cumulative) odds functions and discusses a simple direct optimization method. Section 4 presents our proposed approach based on an EM algorithm in detail, including a two-stage data augmentation, the derivation of the EM algorithm, and the asymptotic variance estimates. Section 5 evaluates the performance of proposed EM algorithm via simulation studies, and Section 6 provides an illustration of our method via a real life application. Some discussions are given in Section 7.

2 Model, data and likelihood

2.1 The PO model

Let T denote the failure time of interest, $\mathbf{x}$ a $p \times 1$ dimensional vector of covariates, and β the corresponding vector of covariate effects. Let $F(t|\mathbf{x})$ and $S(t|\mathbf{x})$ denote the cumulative distribution function (CDF) and survival function of the failure time T given covariate $\mathbf{x}$, respectively.

The PO model specified a proportional relationship in the odds of failure times for different covariate values as follows,

$$\frac{F(t|\mathbf{x})}{1-F(t|\mathbf{x})} = \Lambda_0(t)\exp(\mathbf{x}'\beta), \tag{1}$$

where $\Lambda_0(t) = F_0(t)/\{1-F_0(t)\}$ is the baseline odds function and $F_0(t)$ is the baseline CDF of T given $\mathbf{x}=0$. An equivalent representation of the PO model takes the form

$$\text{logit}\,\{F(t|\mathbf{x})\} = \alpha(t) + \mathbf{x}'\beta, \tag{2}$$

where $\alpha(t) = \log\{\Lambda_0(t)\}$ is the logarithm of baseline odds function and $\text{logit}(y) = \log[y/(1-y)]$ is the so called logit function. Under the PO model, the jth covariate coefficient β_j can be interpreted as the the increase in the log odds of failure by time t attributable to a unit increase in the jth covariate. This interpretation is closely related to that in the logistic regression for binary response, which makes the PO model attractive to non-experts in survival analysis.

Under the PO model, the survival function is $S(t|\mathbf{x}) = 1-F(t|\mathbf{x}) = \{1+\Lambda_0(t)\exp(\mathbf{x}\beta)\}^{-1}$ and the hazard function $\lambda(t|\mathbf{x})$ takes the form

$$\lambda(t|\mathbf{x}) = \frac{\Lambda_0'(t)\exp(\mathbf{x}'\beta)}{1+\Lambda_0(t)\exp(\mathbf{x}'\beta)},$$

where $\Lambda_0'(t)$ is the first derivation of $\Lambda_0(t)$.

2.2 The observed data and likelihood

In this article, we consider right-censored data structure for T. Suppose there are n independent observations. Let T_i and C_i denote the failure time and censoring time for subject i and $\mathbf{x}_i$ the vector of covariates. It is assumed that the failure time and the censoring time are independent conditioning on the covariates. Let $y_i = \min(T_i, C_i)$ be the actual observation and $\delta_i = I(T_i \leq C_i)$ be the censoring indicator, taking 1 if T_i is exactly observed in the study and 0 if T_i is right-censored at C_i for subject i. The observed data are $\{y_i, \delta_i, \mathbf{x}_i\}$ for $i = 1, \cdots, n$, and the observed likelihood function takes the following form

$$L_{obs} = \prod_{i=1}^{n} \lambda(y_i|\mathbf{x}_i)^{\delta_i} S(y_i|\mathbf{x}_i) = \prod_{i=1}^{n} \left[\frac{\Lambda_0'(y_i)\exp(\mathbf{x}_i'\beta)}{1+\Lambda_0(y_i)\exp(\mathbf{x}_i'\beta)} \right]^{\delta_i} \frac{1}{1+\Lambda_0(y_i)\exp(\mathbf{x}_i'\beta)}. \quad (3)$$

3 The proposed method

3.1 Monotone Splines

The observed likelihood (3) contains infinite dimension functions $\Lambda_0(t)$ and $\Lambda_0'(t)$, which contribute greatly to the difficulty of estimation. Acknowledging the fact that $\Lambda_0(t)$ is a nondecreasing function with $\Lambda(0) = 0$, we propose to adopt the monotone splines of (Ramsay 1988) in the following form,

$$\Lambda_0(t) = \sum_{l=1}^{k} \gamma_l b_l(t), \quad (4)$$

where γ_l's are nonnegative spline coefficients and b_l's are integrated spline (I-spline) basis functions. These I-spline basis function are nonnegative piecewise polynomials of specified order $d+1$ or degree d. The degree d largely controls the smoothness of the splines, taking 1 for piecewise linear, 2 for quadratic, and 3 for cubic functions, respectively. Every spline basis function b_l takes 0 in an initial flat region, increases in a middle region, and then plateaus at 1 in the final region. The nonnegative constraint of the spline coefficients ensures that the resulting function is nondecreasing.

Under the monotone spline expression of $\Lambda_0(t)$ in (4), an appealing byproduct is that the derivative of the baseline odds function $\Lambda_0'(t)$ can be approximated directly by

$$\Lambda_0'(t) = \sum_{l=1}^{k} \gamma_l M_l(t), \tag{5}$$

where $M_l(t)$ is the derivative of $b_l(t)$ for $l = 1, \cdots, k$. These monotone spline basis functions b_l's and M_l's are completely determined when the degree and the knot placement are specified and only need to be calculated once. The codes for calculating these basis functions together with our EM algorithm are available upon request.

The spline representations of (4) and (5) are appealing because they are flexible enough to approximate the nondecreasing baseline odds function and its logarithm function naturally with only a finite number of parameters. This allows our approach to enjoy the features of maximum likelihood theory for parametric models while without making strict parametric assumptions on the shape of baseline odds function.

It is worth noting that the approximation here is only valid within a finite interval (a, b), where a and b are usually taken to be (or enclose) the smallest and largest values of y_i's. The number of interior knots taken within (a, b) has a potentially large impact on the modeling flexibility of the target function, although using a small number of interior knots can also provide reasonable results in many cases. For example, Ramsay (1988) recommended to use only a few knots, say, at the median or at the three quartiles. A series of papers in Bayesian survival literature have advocated that using $10 \sim 30$ provides adequate modeling flexibility for various types of data, including Cai et al. (2011), Wang & Dunson (2011),and Lin & Wang (2011) among many others. There are typically two different strategies of knot placements, equally-spaced and quantile-based. Lin et al. (2015) has shown through extensive simulation studies that both strategies work well for their Bayesian method although the equally-spaced strategy outperforms the other with a consistently smaller deviance information criterion (DIC) and a larger log pseudo marginal likelihood (LPML). Based on these literature ideas, we will adopt equally-spaced knots and use model selection criteria such as AIC or BIC to determine the number of knots (McMahan et al. 2013).

3.2 Direct Maximization

With the spline representations of (4) and (5), there are only a finite number of unknown parameters to estimate in the observed likelihood (3). Thus, direct optimization based on the existing optimization package is feasible. Since the spline coefficients are non-negative, a constrained maximization is needed, and specifically we implement the function "L-BFGS-B" in R package "*optm*" for this purpose. The approach for "L-BFGS-B" was based on a quasi-Newton algorithm to solve nonlinear optimization problems with simple bounds on the variables (Byrd et al. 1995). It was found in our simulation that "L-BFGS-B" works well when the right-censoring rate but encountered serious non-convergence problems when the right-censoring rate is high. This problem can be serious since many real life data sets do have a high right-censoring rate. In the following, we present a new approach based on an EM algorithm that provides efficient and robust estimation even when the right-censoring rate is high.

4 Proposed EM algorithm

4.1 A two-stage Data augmentation

The derivation of our EM algorithm is based on a novel data augmentation that contains the following two stages. The first stage of our data augmentation takes advantage of the fact that the survival function under the PO model can be written as the marginal survival function of a frailty PH model with an $\mathcal{E}xp(1)$ for the frailty (Murphy et al. 1997, Shen 1998, McMahan et al. 2013). Specifically we introduce $\phi_i \sim \mathcal{E}xp(1)$ for all $i = 1, \cdots, n$ and $\psi_i \sim \mathcal{E}xp(1)$ only for those exactly observed failures (i.e., $\delta_i = 1$), and the augmented data likelihood takes the following form

$$\mathcal{L}_1 = \prod_{i=1}^{n}\{[\sum_{l=1}^{k}\gamma_l M_l(y_i)]\}^{\delta_i}\exp(\delta_i \mathbf{x}_i'\boldsymbol{\beta})\exp\{-[\sum_{l=1}^{k}\gamma_l b_l(y_i)]exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i+\psi_i\delta_i)\}\exp(-\phi_i-\psi_i\delta_i).$$

Note that all ϕ_i's and ψ_i's are independent in this augmentation. It is clear that integrating out ϕ_i's and ψ_i's in this augmented likelihood $\mathcal{L}_1$ leads to the observed likelihood in (3).

The existence of the additive terms in the augmented likelihood $\mathcal{L}_1$ causes much computational difficulties. In the second stage, we introduce a multinomial latent vector $(v_{i1}, ..., v_{ik}) \sim$ Multinomial$(1, (\frac{1}{k}, ..., \frac{1}{k}))$ if $\delta_i = 1$ in order to turn the additive terms into multiplicative terms in the augmented likelihood. The new augmented likelihood takes the following form

$$\mathcal{L}_C = \prod_{i=1}^{n}\{\prod_{l=1}^{k}[\gamma_l M_l(y_i)]^{v_{il}}\exp(\mathbf{x}_i'\boldsymbol{\beta})\}^{\delta_i}\exp[-\gamma_l M_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i+\psi_i\delta_i)]\}\exp(-\phi_i-\psi_i\delta_i). \tag{6}$$

Integrating out $(v_{i1}, ..., v_{ik})$'s for all $\delta_i = 1$ in this augmented likelihood (6) leads back to the augmented likelihood $\mathcal{L}_1$ up to a constant k. The augmented likelihood (6) will serve as the complete data likelihood for our EM algorithm development, with all ϕ_i's, ψ_i's, and

$(v_{i1}, ..., v_{ik})$'s as missing data.

4.2 Derivation of the EM iteration

Let $\mathcal{D}$ denote the observed data and $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)\prime}, \boldsymbol{\gamma}^{(d)\prime})'$ the current estimate of $\theta = (\beta', \gamma')$ at the dth iteration of the EM algorithm. The E-step of our EM algorithm requires the derivation of the so-called $\mathcal{Q}$ function, which is the expectation of the logarithm of the complete likelihood with respect to the latent variables, condition on the observed data $\mathcal{D}$ and the current parameter estimate $\boldsymbol{\theta}^{(d)}$. Here the $\mathcal{Q}$ function takes the form

$$\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(d)}) = \sum_{i=1}^{n}\{[\sum_{l=1}^{k}(\log(\gamma_l M_l(y_i))E(v_{il}) + \mathbf{x}_i'\boldsymbol{\beta}]\delta_i - \sum_{l=1}^{k}\gamma_l b_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})E(\phi_i + \psi_i\delta_i) - E(\phi_i + \psi_i\delta_i)\}.$$

All the expectations in the above expression are taken conditionally on the observed data, and they have closed form as follows,

$$E(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = [\sum_{l=1}^{k}\gamma_l^{(d)} b_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1]^{-1}, \tag{7}$$

$$E(\psi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = [\sum_{l=1}^{k}\gamma_l^{(d)} b_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1]^{-1}, \tag{8}$$

$$E(v_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \frac{\gamma_l^{(d)} M_l(y_i)}{\sum_{l=1}^{k}\gamma_l^{(d)} M_l(y_i)}, \quad \text{for } l = 1, \cdots, k. \tag{9}$$

These expectations are obtained based on the following facts that given the observed data the conditional distributions of ϕ_i and ψ_i are exponentially distributed and $v_i = (v_{i1}, \cdots, v_{ik})$ is a multinomial distribution. Note that ψ_i and $v_i = (v_{i1}, \cdots, v_{ik})$ are introduced only when $\delta_i = 1$.

The M-step in our EM algorithm seeks to find the maximizer of the $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(d)})$. To this end, we fist consider the first partial derivatives of $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(d)})$ w.r.t $\boldsymbol{\theta}$ and set them equal to

0. This leads to the following equations

$$\begin{aligned}\frac{\partial \mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(d)})}{\partial \beta} &= \sum_{i=1}^{n}\left\{\delta_i - \sum_{l=1}^{k}\gamma_l b_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})E(\phi_i+\psi_i\delta_i)\right\}\mathbf{x}_i = 0\\ \frac{\partial \mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(d)})}{\partial \gamma_l} &= \sum_{i=1}^{n}\left\{\gamma_l^{-1}E(v_{il})\delta_i - b_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})E(\phi_i+\psi_i\delta_i)\right\} = 0, \quad \forall l.\end{aligned}$$

It is clear that the equation $\partial\mathcal{Q}/\partial\gamma_l = 0$ has an explicit solution of γ_l,

$$\boldsymbol{\gamma_l^*(\beta) = \frac{\sum_{i=1}^{n} E(v_{il})\delta_i}{\sum_{i=1}^{n} b_l(y_i)\exp(\mathbf{x}_i'\beta)E(\phi_i+\psi_i\delta_i)}}, \tag{10}$$

as a function of β for each $l = 1, \cdots, k$. Thus, one can plug these solutions $\gamma_l^*(\beta)$ in (10) into equation $\partial\mathcal{Q}/\partial\beta = 0$ and solve it for $\beta^{(d+1)}$. Then we can obtain $\gamma_l^{(d+1)}$ by updating the expression (10) using $\beta^{(d+1)}$ for $l = 1, \cdots, k$. It can be proved that $\theta^{(d+1)} = (\beta^{(d+1)'}, \gamma^{(d+1)'})'$ is the unique global maximizer of $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(d)})$. See the sketched proof in the supplementary file.

Our EM algorithm is summarized as follows. First set $d = 0$ and initialize $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \gamma^{(d)'})'$ and then repeat the following two steps until convergence:

1. Obtain $\boldsymbol{\beta}^{(d+1)}$ by solving the following system of p equations

$$\sum_{i=1}^{n}\left\{\delta_i - \sum_{l=1}^{k}\gamma_l^*(\beta)b_l(y_i)\exp(\mathbf{x}_i'\beta)E(\phi_i+\psi_i\delta_i)\right\}\mathbf{x}_i = 0,$$

2. Obtain $\gamma_l^{(d+1)}$ through explicit updating

$$\gamma_l^{(d+1)} = \frac{\sum_{i=1}^{n} E(v_{il})\delta_i}{\sum_{i=1}^{n} b_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d+1)})E(\phi_i+\psi_i\delta_i)},$$

for $l = 1, \cdots, k$ and increase d by 1.

This EM algorithm is straightforward to implement as it only involves solving a low-dimensional system of equations for the regression coefficients and then updating the spline coefficients in simple explicit form at each iteration. It is worthy noting that the explicit form of γ_l's automatically satisfies the non-negativity constraints, which adds great efficiency of this algorithm.

4.3 Variance Estimation

Let $\hat{\theta}$ denote the converged value from the EM sequence. It is straightforward to show that $\hat{\theta}$ is indeed the root of the score equation based on the observed likelihood (3). It is important to obtain the variance estimate $\text{var}(\hat{\beta})$ of the regression parameter estimates. Since there are only a finite number of parameters through the spline approximation in the model, the conventional likelihood theory can be applied here. Thus, under regularity conditions, $\hat{\theta}$ has an asymptotically normal distribution with mean θ and variance $\{I(\hat{\theta})\}^{-1}$, where $I(\hat{\theta})$ is the observed information matrix evaluated at $\hat{\theta}$. To obtain this variance estimate, we adopt Louis's method (Louis 1982) to quantify $I(\theta)$ as follows,

$$I(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{var}\{\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} | \mathcal{D}, \hat{\boldsymbol{\theta}}\}.$$

All the quantities in $\text{var}\{\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} | \mathcal{D}, \hat{\boldsymbol{\theta}}\}$ and $\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ have explicit forms as presented in Appendix A, making the variance calculation straightforward. Wald inferences can be made to assess the covariate effects. Various functions, such as the baseline CDF, survival function, and odds function, can be estimated directly based on the EM output. The R function of this algorithm is available upon request.

5 Simulation Studies

Extensive simulation studies were conducted to evaluate the performance of the proposed method. We considered a general PO model with both continuous and discrete covariates: $x_1 \sim \mathcal{N}(0,1)$ and $x_2 \sim Bernoulli(0.5)$. The true baseline odds function $\Lambda_0(t)$ was taken to be $\Lambda_0(t) = \log(1+t)+t^{1.5}$ and $\Lambda_0(t) = \log(1+t)+t^3+sin(t)$, and both regression parameters took values on -1,0, and 1 respectively. The censoring time C for each subject was generated independently from an exponential distribution with rate parameter τ. Different values of τ determine the rate of right-censoring. We considered three different scenarios of low, medium, and high censoring rates by taking τ to be 0.1, 1 and 5, respectively. The right-censoring rate varies between 4.5% and 28.5% across all parameter configurations in the low censoring rate scenario (Scenario I), between 31% and 69% in the medium censoring rate scenario (Scenario II), and between 60% and 93.5% in the high censoring rate scenario (Scenario III), respectively.

For each parameter configuration, we generated 500 independent data sets with sample size $n = 200$. For the spline specifications, we took 9 equally spaced interior knots between 0 and maximum of the observation times in each data set and used 3 for the degree of the I-spline to ensure adequate smoothness for the baseline odds function. Both direct maximization and the proposed EM algorithm were applied for each data set. The initial values were specified to be 1 for all the parameters under both methods. Random initial values from Uniform $\mathcal{U}_{(0,1)}$ were also tried and worked well for our EM algorithm. It should be noted that the direct maximization encountered serious non-convergence problems in Scenario III, where the right-censoring rates are high. Thus, only the results from EM algorithm are summarized in scenario III.

Tables 1 and 2 summarize the estimation results on the regression parameters from the proposed EM algorithm and direct optimization method (DM) in Scenarios I and II respectively, and Table 3 presents the results from the EM algorithm in Scenario III. These results include BIAS, the difference between the average of 500 parameter estimates and the true value, ESE, the average of the 500 estimated standard errors, SSD, the sample standard deviation of the 500 parameter estimates, and CP95, the 95% coverage probability based on the Wald confidence intervals, for all parameters in each setup. From these results, it is clear that the proposed EM algorithm has a good performance in estimating the regression parameters. Firstly, the biases of our estimates are all quite small, indicating the asymptotic unbiasedness of our estimates. Secondly, the ESDs are close to the corresponding SSDs, suggesting that the variance estimates based on Louis's method are accurate. Lastly, the empirical coverage probabilities CP95 are close to the nominal value 0.95, suggesting that the asymptotic normality distributions of our estimates are valid. Direct optimization method also provides good and comparable results as the proposed EM algorithm in Scenarios I and II as seen in Tables 1 and 2. However, the method failed to work for the high censoring rate scenario.

In order to evaluate the performance of our proposed methods in estimating the baseline CDF (survival or odds) function, we summarized the mean and maximum of local mean squared errors (MSE) of the baseline CDF estimates at some pre-specified grid points. The local MSE of $\hat{F}_0(t)$ is calculated as follows,

$$\mathrm{MSE}\{\hat{F}_0(t)\} = \frac{1}{M}\sum_{j=1}^{M}\{\hat{F}_0^{(j)}(t) - F_0(t)\}^2,$$

where $F_0(t)$ is the true baseline CDF and $\hat{F}_0^{(j)}(t)$ is the estimated baseline CDF from the jth data set at time t. Table 4 presents the mean and maximum of the local MSEs for the baseline CDF (F_0) estimates in all scenarios. The grid points were taken to be 50 evenly-spaced points between 0 and a constant c, which was the median of the 500 maximums of the observed values from the 500 data sets for each parameter configuration. The c value varied between 10.00 and 15.74, 3.99 and 5.72 in Scenario I under two different baseline odds functions, between 2.10 and 2.75 in Scenario and II, and between 0.57 and 0.63 in Scenario III across all parameter configurations. As seen clearly in Table 4, all of the mean and maximum of the local MSEs are quite small, indicating that the proposed methods provide accurate estimation on the baseline CDF (survival or odds) as well.

All the simulations were conducted using R on a computer with a 3.60 GHz processor and 32.0 GB of memory. On average, it took 0.1 second for the proposed EM algorithm to converge and return the variance estimates for each data set. It took 0.4 second per data set for the direct optimization to return the result in the cases of low and medium right-censoring rate.

Table 1: Simulation results of the estimated regression parameters using the proposed EM algorithm (EM) and direct optimization (DM) for scenario I.

				$\Lambda_0(t) = \log(1+t) + t^{1.5}$				$\Lambda_0(t) = \log(1+t) + t^3 + sin(t)$			
β_1	β_2	Method	Est	BIAS	ESD	SSD	CP95	BIAS	ESD	SSD	CP95
1	1	EM	$\hat{\beta}_1$	-0.002	0.140	0.145	0.942	0.011	0.141	0.147	0.954
			$\hat{\beta}_2$	0.004	0.258	0.267	0.936	0.025	0.260	0.254	0.948
		DM	$\hat{\beta}_1$	-0.002	0.140	0.145	0.944	0.008	0.141	0.146	0.954
			$\hat{\beta}_2$	0.006	0.258	0.268	0.942	0.028	0.260	0.255	0.952
	0	EM	$\hat{\beta}_1$	-0.001	0.141	0.147	0.946	0.011	0.143	0.144	0.942
			$\hat{\beta}_2$	-0.009	0.252	0.268	0.942	-0.013	0.254	0.266	0.926
		DM	$\hat{\beta}_1$	-0.003	0.141	0.146	0.952	0.006	0.145	0.143	0.948
			$\hat{\beta}_2$	-0.006	0.252	0.268	0.946	-0.005	0.260	0.267	0.920
	-1	EM	$\hat{\beta}_1$	0.006	0.143	0.145	0.964	0.024	0.146	0.144	0.952
			$\hat{\beta}_2$	-0.009	0.262	0.254	0.956	-0.031	0.268	0.281	0.954
		DM	$\hat{\beta}_1$	0.003	0.144	0.144	0.960	0.018	0.147	0.143	0.942
			$\hat{\beta}_1$	0.015	0.263	0.253	0.962	-0.020	0.272	0.287	0.950
0	1	EM	$\hat{\beta}_1$	0.001	0.125	0.123	0.954	0.008	0.126	0.129	0.948
			$\hat{\beta}_2$	0.012	0.258	0.260	0.950	0.004	0.260	0.266	0.944
		DM	$\hat{\beta}_1$	0.001	0.125	0.123	0.954	0.008	0.126	0.128	0.950
			$\hat{\beta}_2$	0.015	0.258	0.260	0.958	0.008	0.259	0.267	0.940
	0	EM	$\hat{\beta}_1$	0.011	0.126	0.126	0.950	0.007	0.127	0.130	0.940
			$\hat{\beta}_2$	-0.001	0.252	0.244	0.956	0.019	0.252	0.251	0.952
		DM	$\hat{\beta}_1$	0.012	0.126	0.125	0.950	0.008	0.127	0.130	0.940
			$\hat{\beta}_1$	0.004	0.251	0.246	0.958	0.024	0.254	0.254	0.948
	-1	EM	$\hat{\beta}_1$	0.002	0.127	0.132	0.944	-0.001	0.129	0.129	0.954
			$\hat{\beta}_2$	-0.004	0.261	0.256	0.966	-0.030	0.267	0.269	0.940
		DM	$\hat{\beta}_1$	0.012	0.127	0.131	0.940	-0.001	0.128	0.128	0.956
			$\hat{\beta}_2$	0.009	0.261	0.256	0.964	-0.024	0.269	0.273	0.950
-1	1	EM	$\hat{\beta}_1$	-0.003	0.140	0.141	0.956	-0.024	0.141	0.149	0.936
			$\hat{\beta}_2$	0.006	0.258	0.266	0.954	0.014	0.259	0.259	0.948
		DM	$\hat{\beta}_1$	-0.001	0.134	0.140	0.952	-0.019	0.142	0.147	0.948
			$\hat{\beta}_2$	0.008	0.258	0.266	0.952	0.017	0.261	0.259	0.952
	0	EM	$\hat{\beta}_1$	-0.002	0.142	0.149	0.948	-0.014	0.143	0.148	0.934
			$\hat{\beta}_2$	-0.001	0.253	0.246	0.960	-0.022	0.254	0.249	0.958
		DM	$\hat{\beta}_1$	0.001	0.142	0.148	0.944	-0.012	0.143	0.148	0.946
			$\hat{\beta}_2$	0.003	0.252	0.245	0.960	-0.018	0.256	0.252	0.960
	-1	EM	$\hat{\beta}_1$	-0.011	0.144	0.146	0.940	-0.008	0.145	0.156	0.926
			$\hat{\beta}_2$	-0.001	0.264	0.274	0.950	-0.013	0.264	0.269	0.938
		DM	$\hat{\beta}_1$	-0.009	0.144	0.146	0.944	-0.002	0.146	0.154	0.928
			$\hat{\beta}_2$	0.006	0.262	0.275	0.942	0.001	0.273	0.270	0.940

Table 2: Simulation results of the estimated regression parameters using the proposed EM algorithm (EM) and direct optimization (DM) for scenario II.

				$\Lambda_0(t) = \log(1+t) + t^{1.5}$				$\Lambda_0(t) = \log(1+t) + t^3 + sin(t)$			
β_1	β_2	Method	Est	BIAS	ESD	SSD	CP95	BIAS	ESD	SSD	CP95
1	1	EM	$\hat{\beta}_1$	0.015	0.158	0.160	0.946	0.012	0.153	0.156	0.960
			$\hat{\beta}_2$	0.012	0.285	0.286	0.946	0.014	0.278	0.284	0.936
		DM	$\hat{\beta}_1$	0.013	0.157	0.162	0.944	0.009	0.152	0.154	0.958
			$\hat{\beta}_2$	0.014	0.284	0.287	0.954	0.017	0.277	0.285	0.936
	0	EM	$\hat{\beta}_1$	0.027	0.166	0.169	0.956	0.014	0.159	0.161	0.948
			$\hat{\beta}_2$	-0.017	0.289	0.283	0.956	0.008	0.279	0.297	0.930
		DM	$\hat{\beta}_1$	0.025	0.165	0.169	0.952	0.012	0.158	0.160	0.952
			$\hat{\beta}_2$	-0.016	0.288	0.285	0.954	0.010	0.278	0.299	0.924
	-1	EM	$\hat{\beta}_1$	0.014	0.177	0.172	0.952	0.024	0.167	0.178	0.944
			$\hat{\beta}_2$	-0.020	0.319	0.326	0.950	-0.018	0.304	0.312	0.936
		DM	$\hat{\beta}_1$	0.010	0.174	0.172	0.952	0.019	0.166	0.176	0.944
			$\hat{\beta}_2$	-0.018	0.315	0.324	0.956	-0.008	0.300	0.309	0.946
0	1	EM	$\hat{\beta}_1$	-0.006	0.137	0.145	0.934	0.000	0.133	0.135	0.946
			$\hat{\beta}_2$	0.016	0.280	0.280	0.946	0.025	0.274	0.279	0.950
		DM	$\hat{\beta}_1$	-0.007	0.137	0.146	0.938	0.000	0.133	0.135	0.940
			$\hat{\beta}_2$	0.017	0.280	0.279	0.950	0.030	0.273	0.279	0.948
	0	EM	$\hat{\beta}_1$	0.002	0.142	0.147	0.944	-0.010	0.138	0.143	0.934
			$\hat{\beta}_2$	-0.010	0.283	0.292	0.950	0.003	0.272	0.279	0.956
		DM	$\hat{\beta}_1$	-0.003	0.142	0.146	0.944	-0.010	0.138	0.143	0.932
			$\hat{\beta}_2$	0.007	0.283	0.291	0.956	0.008	0.273	0.279	0.964
	-1	EM	$\hat{\beta}_1$	0.004	0.151	0.153	0.946	0.002	0.145	0.142	0.956
			$\hat{\beta}_2$	-0.067	0.312	0.335	0.942	-0.043	0.300	0.320	0.926
		DM	$\hat{\beta}_1$	0.004	0.151	0.153	0.946	0.002	0.145	0.141	0.958
			$\hat{\beta}_2$	-0.064	0.312	0.333	0.946	-0.034	0.297	0.320	0.936
-1	1	EM	$\hat{\beta}_1$	-0.028	0.159	0.164	0.942	-0.005	0.153	0.156	0.946
			$\hat{\beta}_2$	0.035	0.286	0.306	0.932	0.034	0.277	0.278	0.960
		DM	$\hat{\beta}_1$	-0.025	0.158	0.163	0.942	0.000	0.152	0.154	0.950
			$\hat{\beta}_2$	0.036	0.285	0.306	0.932	0.036	0.277	0.278	0.952
	0	EM	$\hat{\beta}_1$	-0.026	0.166	0.166	0.952	-0.016	0.158	0.162	0.950
			$\hat{\beta}_2$	-0.019	0.292	0.279	0.958	-0.033	0.279	0.287	0.940
		DM	$\hat{\beta}_1$	-0.026	0.165	0.166	0.952	-0.011	0.158	0.160	0.948
			$\hat{\beta}_2$	0.027	0.289	0.280	0.956	-0.027	0.278	0.286	0.942
	-1	EM	$\hat{\beta}_1$	-0.030	0.176	0.183	0.944	-0.026	0.168	0.173	0.950
			$\hat{\beta}_2$	-0.032	0.318	0.315	0.958	-0.021	0.306	0.326	0.936
		DM	$\hat{\beta}_1$	-0.028	0.175	0.182	0.944	-0.021	0.167	0.171	0.948
			$\hat{\beta}_2$	0.027	0.315	0.315	0.960	-0.009	0.301	0.327	0.946

Table 3: Simulation results of the estimated regression parameters from EM algorithm for scenario III.

			$\Lambda_0(t) = \log(1+t) + t^{1.5}$				$\Lambda_0(t) = \log(1+t) + t^3 + sin(t)$			
β_1	β_2	Est	BIAS	ESD	SSD	CP95	BIAS	ESD	SSD	CP95
1	1	$\hat{\beta}_1$	0.044	0.199	0.203	0.944	0.020	0.185	0.179	0.958
		$\hat{\beta}_2$	0.039	0.360	0.345	0.962	0.004	0.336	0.332	0.954
	0	$\hat{\beta}_1$	0.036	0.220	0.225	0.956	0.025	0.203	0.208	0.950
		$\hat{\beta}_2$	-0.021	0.386	0.376	0.960	-0.005	0.357	0.369	0.944
	-1	$\hat{\beta}_1$	0.051	0.251	0.245	0.970	0.033	0.226	0.232	0.954
		$\hat{\beta}_2$	-0.054	0.461	0.500	0.946	-0.023	0.413	0.435	0.944
0	1	$\hat{\beta}_1$	-0.003	0.174	0.180	0.942	0.010	0.161	0.169	0.944
		$\hat{\beta}_2$	0.038	0.364	0.361	0.952	0.017	0.332	0.354	0.948
	0	$\hat{\beta}_1$	0.005	0.196	0.196	0.960	-0.001	0.179	0.188	0.948
		$\hat{\beta}_2$	0.016	0.396	0.391	0.958	0.001	0.351	0.360	0.936
	-1	$\hat{\beta}_1$	-0.034	0.226	0.219	0.960	0.004	0.203	0.199	0.958
		$\hat{\beta}_2$	-0.055	0.492	0.512	0.950	0.002	0.427	0.438	0.932
-1	1	$\hat{\beta}_1$	-0.028	0.199	0.202	0.952	-0.024	0.185	0.195	0.944
		$\hat{\beta}_2$	0.021	0.361	0.355	0.954	0.071	0.338	0.345	0.946
	0	$\hat{\beta}_1$	-0.028	0.222	0.232	0.948	-0.045	0.205	0.212	0.952
		$\hat{\beta}_2$	-0.001	0.387	0.383	0.952	-0.013	0.360	0.366	0.946
	-1	$\hat{\beta}_1$	-0.048	0.249	0.268	0.948	-0.037	0.227	0.227	0.960
		$\hat{\beta}_2$	-0.037	0.458	0.460	0.956	-0.043	0.417	0.428	0.954

Table 4: Mean and maximum of local MSEs of the baseline CDF estimate $\hat{F}_0(t)$ obtained by using the proposed EM algorithm (EM) and direct optimization (DM) in all simulation scenarios. The true baseline odds function taking $\Lambda_0(t) = \log(1+t) + t^{1.5}(B_1)$ and $\Lambda_0(t) = \log(1+t) + t^3 + sin(t)(B_2)$.

			Scenario I		Scenario II		Scenario III	
$\Lambda_0(t)$	(β_1, β_2)	Method	meanMSE	maxMSE	meanMSE	maxMSE	meanMSE	maxMSE
B_1	(1,1)	EM	0.0007	0.0022	0.0033	0.0042	0.0044	0.0121
		DM	0.0007	0.0022	0.0033	0.0042	-	-
	(1,0)	EM	0.0005	0.0020	0.0030	0.0040	0.0045	0.0112
		DM	0.0005	0.0020	0.0030	0.0040	-	-
	(1,-1)	EM	0.0005	0.0022	0.0029	0.0034	0.0050	0.0130
		DM	0.0005	0.0022	0.0029	0.0034	-	-
	(0,1)	EM	0.0008	0.0019	0.0029	0.0043	0.0044	0.0124
		DM	0.0008	0.0019	0.0029	0.0043	-	-
	(0,0)	EM	0.0007	0.0020	0.0027	0.0036	0.0043	0.0115
		DM	0.0007	0.0020	0.0027	0.0036	-	-
	(0,-1)	EM	0.0005	0.0020	0.0028	0.0034	0.0048	0.0121
		DM	0.0005	0.0020	0.0028	0.0034	-	-
	(-1,1)	EM	0.0007	0.0022	0.0036	0.0047	0.0046	0.0121
		DM	0.0007	0.0022	0.0036	0.0047	-	-
	(-1,0)	EM	0.0006	0.0021	0.0031	0.0037	0.0051	0.0121
		DM	0.0006	0.0021	0.0031	0.0037	-	-
	(-1,-1)	EM	0.0005	0.0022	0.0028	0.0034	0.0047	0.0112
		DM	0.0005	0.0022	0.0028	0.0034	-	-
B_2	(1,1)	EM	0.0008	0.0023	0.0025	0.0032	0.0048	0.0123
		DM	0.0008	0.0023	0.0025	0.0032	-	-
	(1,0)	EM	0.0007	0.0024	0.0023	0.0033	0.0050	0.0118
		DM	0.0007	0.0024	0.0023	0.0033	-	-
	(1,-1)	EM	0.0006	0.0025	0.0021	0.0035	0.0055	0.0122
		DM	0.0006	0.0025	0.0021	0.0035	-	-
	(0,1)	EM	0.0009	0.0021	0.0025	0.0030	0.0053	0.0133
		DM	0.0009	0.0021	0.0025	0.0030	-	-
	(0,0)	EM	0.0008	0.0022	0.0021	0.0028	0.0048	0.0109
		DM	0.0008	0.0022	0.0021	0.0028	-	-
	(0,-1)	EM	0.0007	0.0023	0.0018	0.0027	0.0054	0.0126
		DM	0.0007	0.0023	0.0018	0.0027	-	-
	(-1,1)	EM	0.0009	0.0025	0.0024	0.0029	0.0049	0.0119
		DM	0.0009	0.0025	0.0024	0.0029	-	-
	(-1,0)	EM	0.0007	0.0027	0.0023	0.0032	0.0049	0.0113
		DM	0.0007	0.0027	0.0023	0.0032	-	-
	(-1,-1)	EM	0.0007	0.0026	0.0022	0.0037	0.0053	0.0119
		DM	0.0007	0.0026	0.0022	0.0037	-	-

6 Data Applications

Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial is a randomized, controlled trial that is sponsored by the United States National Cancer Institute since November 1993. Eligible participants aged between 55 and 74, had no history of any PLCO cancer, and did not participated in any other cancer screening or primary prevention trials at the enrollment.

As a part of the PLCO study, the prostate cancer screening data were collected on male participants in the intervention arm who underwent blood test for prostate-specific antigen (PSA) analysis at enrollment and then annually for the next 5 years. The primary goal of our analysis is to assess the association of risk factors with the age at prostate cancer diagnosis. The dataset used here was released by National Cancer Institute in 2011 and had 34,175 participants in total. Our data analysis is based on 33,230 participants after deleting some participants with missing values in some covariates. Among the 33,230 observations, 3,362 were diagnosed to have prostate cancer, and the others had not contracted prostate cancer yet at their 5th year follow up, so their ages at diagnosis were essentially right-censored. Thus, our response variable has a right-censored data structure with an extremely high censoring rate, nearly 90%.

We considered the following covariates for our analysis: education, binary with 1 indicating a college education; race, with three categories Caucasian, African American, and other; obesity, binary with 1 indicating obesity; heart, binary with 1 indicating presence of heart disease; stroke, binary with 1 indicating a previous stroke; diabetes, binary with 1 indicating diabetic; colitis, binary with 1 indicating a positive status; hepatitis, binary with

1 indicating a positive status; aspirin, binary with 1 indicating regular use; ibuprofen, binary with 1 indicating regular use; and family history, binary with 1 indicating that an immediate relative had prostate cancer.

Direct maximization failed to work for this analysis due to the extremely high censoring rate and the large sample size. We applied the proposed EM algorithm to this data set with different numbers of equally-spaced knots and degree for the spline specifications. Then we conduct model selection to determine the number of interior knots and degree based on the AIC criteria. It was observed that using different values for the number of knots and degree yielded very close estimation results as shown in Table 1 of the supplementary file. This phenomenon has also been reported in many literature research using monotone splines, such as Wang & Dunson (2011) and Lin & Wang (2011) among others.

The model using 12 equally spaced interior knots and degree 2 yielded the smallest AIC value as seen in Table 1 of the supplementary file, and the corresponding results are summarized in Table 5. As seen in Table 5, all covariates are significantly associated with age of diagnosis of prostate cancer except colitis and the use of Aspirin. Black race has the highest risk of developing prostate cancer among all race groups, followed by white race. Family history, college education, obesity, and use of ibuprofen are all positively associated with prostate cancer, while heart disease, stroke, diabetes, and hepatitis seem to have a protective effect on prostate cancer.

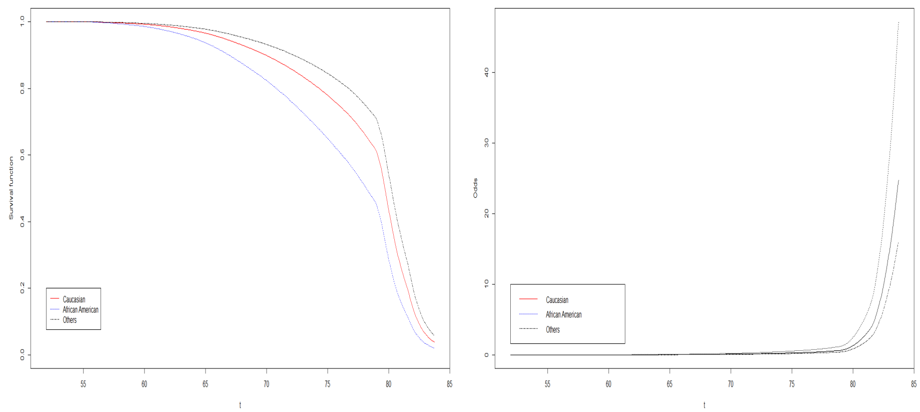

Figure 1: The estimated survival functions and odds functions for participants in three different race groups holding all other covariates at the baseline values.

Table 5: Results of the PLCO data analysis in terms of the point estimate (Point), standard error (SE), and p-value from the proposed EM algorithm using 12 equally spaced interior knots and degree 2.

Covariate	Point	SE	P-value
Family history	0.5094	0.0632	0.0000
Black Race	0.6461	0.0837	0.0000
Other Race	-0.4322	0.0922	0.0000
Education	0.1505	0.0400	0.0002
Obesity	0.1702	0.0475	0.0003
Aspirin	-0.0575	0.0400	0.1499
Ibuprofen	0.1566	0.0455	0.0006
Heart Disease	-0.3809	0.0606	0.0000
Stroke	-0.3094	0.1262	0.0143
Diabetes	-0.5678	0.0772	0.0000
Hepatitis	-0.2144	0.1086	0.0483
Colitis	-0.3206	0.2069	0.1213

Figure 1 shows the estimated survival functions and odds functions for participants in three race groups: African American, Caucasian, and other race, with all other covariates taking the baseline values. These plots indicate a clear difference in the survival and odds functions among the three race groups.

7 Concluding remarks

In this paper, we propose new approaches to analyze right-censored data under the PO model. By adopting monotone splines of Ramsay (1988), we can approximate both the baseline odds function and its derivative through I splines and M splines respectively, and this reduces the number of parameters to be estimated to a finite number. Direct optimization is applicable and has a good performance when the right censoring rate is low and moderate but fails to work when the right censoring rate is high as shown in our simulation and data analysis. To remedy this, we further proposed an EM algorithm based on a two-stage data augmentation. Our EM algorithm is easy to implement, robust to initial values, and fast to converge, and variance estimates are also provided in closed form based on the EM output. The EM algorithm has excellent performance in estimating both the regression parameters and survival functions even when the right-censoring rate is very high as seen in our simulation and data analysis. Due to its excellent estimation performance and easy implementation, this new approach based on the EM algorithm is expected to be widely used for analyzing right-censored data and will make the PO model more appealing as an alternative to the PH model to both statisticians and non-statisticians.

Appendix A. Quantities needed for calculating $\mathrm{var}(\hat{\boldsymbol{\theta}})$

All the necessary entries in $\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ are

$$\begin{aligned}
\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} &= -\sum_{i=1}^{n}\sum_{l=1}^{k} \gamma_l b_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})E(\phi_i + \psi_i\delta_i)\mathbf{x}_i\mathbf{x}_i',\\
\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial\gamma_l\partial\boldsymbol{\beta}'} &= -\sum_{i=1}^{n} b_l(y_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})E(\phi_i + \psi_i\delta_i)]\mathbf{x}_i',\\
\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial\gamma_l^2} &= -\gamma_l^{-2}\sum_{i=1}^{n} E(v_{il})\delta_i,\\
\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial\gamma_l\partial\gamma_{l'}} &= 0, \qquad l \neq l'.
\end{aligned}$$

The necessary entries of $\mathrm{var}\{\partial \log \mathcal{L}_C(\theta)/\partial\theta|\mathcal{D}, \hat{\theta}\}$ are given by:

$$\begin{aligned}
\mathrm{var}\left(\frac{\partial \log(\mathcal{L}_C)}{\partial\boldsymbol{\beta}}\right) &= \sum_{i=1}^{n}\{\Lambda_0(y_i)\}^2\exp(2\mathbf{x}_i'\boldsymbol{\beta})\mathrm{var}(\phi_i + \psi_i\delta_i)\mathbf{x}_i\mathbf{x}_i',\\
\mathrm{var}\left(\frac{\partial \log(\mathcal{L}_C)}{\partial\gamma_l}\right) &= \sum_{i=1}^{n}\left[\frac{\delta_i M_l(y_i)}{\gamma_l\Lambda_0'(y_i)}\{1 - \frac{\gamma_l M_l(y_i)}{\Lambda_0'(y_i)}\} + b_l(y_i)b_{l'}(y_i)\exp(2\mathbf{x}_i'\beta)\mathrm{var}(\phi_i + \psi_i\delta_i)\right],\\
\mathrm{cov}(\frac{\partial \log(\mathcal{L}_C)}{\partial\gamma_l}, \frac{\partial \log(\mathcal{L}_C)}{\partial\gamma_{l'}}) &= \sum_{i=1}^{n}\{b_l(y_i)b_{l'}(y_i)\exp(2\mathbf{x}_i'\boldsymbol{\beta})\mathrm{var}(\phi_i + \psi_i\delta_i) - \frac{\delta_i M_l(y_i)M_{l'}(y_i)}{[\Lambda_0'(t_i)]^2}\}, \quad l \neq l',\\
\mathrm{cov}(\frac{\partial \log(\mathcal{L}_C)}{\partial\boldsymbol{\beta}}, \frac{\partial \log(\mathcal{L}_C)}{\partial\gamma_l}) &= \sum_{i=1}^{n} b_l(y_i)\Lambda_0(y_i)\exp(2\mathbf{x}_i'\boldsymbol{\beta})\mathrm{var}(\phi_i + \psi_i\delta_i)\mathbf{x}_i, \forall l,
\end{aligned}$$

where $\mathrm{var}(\phi_i + \psi_i\delta_i) = (1+\delta_i)\{\sum_{l=1}^{k}\gamma_l b_l(t_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}) + 1\}^{-2}$.

References

Banerjee, S. & Dey, K. D. (2005), 'Semiparametric proportional odds models for spatially correlated survival data', *Lifetime Data Analysis* **11**, 175–191.

Bennett, S. (1983a), 'Analysis of survival data by the proportional odds model', *Statistics in Medicine* **2**, 273–277.

Bennett, S. (1983b), 'Log-logistic regression models for survival data', *Applied Statistics* **32**, 165–171.

Byrd, X., Lu, P., Nocedal, J. & Zhu, C. (1995), 'A limited memory algorithm for bound constrained optimization', *SIAM Journal on Scientific Computing* **16**, 1190–1208.

Cai, B., Lin, X. & Wang, L. (2011), 'Bayesian proportional hazards model for current status data with monotone splines', *Computational Statistical Data Analysis* **55**, 2644–2651.

Hanson, T. & Yang, M. (2007), 'Bayesian semiparametric proportional odds models', *Biometrics* **63**, 88–95.

Huang, J. & Rossini, J. (1997), 'Sieve estimation for the proportional odds model with interval-censoring', *Journal of American Statistical Association* **92**, 960–967.

Lin, X. & Wang, L. (2011), 'Bayesian proportional odds models for analyzing current status data: Univariate, clustered, and multivariate', *Communications in Statistics - Simulation and Computation* **40**, 1171–1181.

Louis, T. (1982), 'Finding the observed information matrix when using the em algorithm', *Journal of the Royal Statistical Society* **44**, 226–233.

McCullagh, P. (1977), 'A logistic model for paired comparisons with ordered categorical data', *Biometrika* **64**, 449–53.

McCullagh, P. (1980), 'Regression models for ordinal data', *Journal of the Royal Statistical Society* **42**, 109–142.

McMahan, C., Wang, L. & Tebbs, J. (2013), 'Regression analysis for current status data using the em algorithm', *Statistics in Medicine* **32**, 4452–4466.

Murphy, S., Rossini, A. & van der Vaart, A. W. (1997), 'Maximum likelihood estimation in the proportional odds model', *Journal of the American Statistical Association* **92**, 968–976.

Pettitt, A. (1984), 'Proportional odds models for surval data and estimates using ranks', *Journal of the Royal Statistical Society* **33**, 169–175.

Ramsay, J. O. (1988), 'Monotone regression splines in action', *Statistical Science* **3**, 425–461.

Robinowitz, D., Betensky, R. A. & Tsiatis, A. A. (2000), 'Using conditional logistic regression to fit proportional odds models to interval censored data', *Biometrics* **56**, 511–518.

Rossini, A. J. & Tsiatis, A. A. (1996), 'A semiparametric proportional odds regression model for the analysis of current status data', *Journal of the American Statistical Association* **91**, 713–721.

Royston, P. & Parmar, M. K. B. (2002), 'Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects', *Statistics in Medicine* **21**, 2175–2197.

Shen, X. T. (1998), 'Proportional odds regression and sieve maximum likelihood estimation', *Biometrika* **85**, 165–177.

Wang, L. & Dunson, D. (2011), 'Semiparametric bayes proportional odds models for current status data with underreporting', *Biometrics* **67**, 1111–1118.