AISM (2006) 58: 511–526
DOI 10.1007/s10463-005-0026-4

Shuangge Ma · Michael R. Kosorok

Adaptive penalized M-estimation with current status data

Received: 18 August 2004 / Revised: 7 February 2005 / Published online: 11 July 2006

Abstract Current status data arises when a continuous response is reduced to an indicator of whether the response is greater or less than a random threshold value. In this article we consider adaptive penalized M-estimators (including the penalized least squares estimators and the penalized maximum likelihood estimators) for nonparametric and semiparametric models with current status data, under the assumption that the unknown nonparametric parameters belong to unknown Sobolev spaces. The Cox model is used as a representative of the semiparametric models. It is shown that the modified penalized M-estimators of the nonparametric parameters can achieve adaptive convergence rates, even when the degrees of smoothing are not known in advance. $\sqrt{n}$ consistency, asymptotic normality and inference based on the weighted bootstrap for the estimators of the regression parameter in the Cox model are also established. A simulation study is conducted for the Cox model to evaluate the finite sample efficacy of the proposed approach and to compare it with the ordinary maximum likelihood estimator. It is demonstrated that the proposed method is computationally superior.We apply the proposed approach to the California Partner Study analysis.

Keywords Adaptive semiparametric estimation · Current status data · Penalized M-estimator

1 Introduction

Current status data, which is also known as case I interval censored data, arises in studies in which the target measurement is the time of occurrence of some

S. Ma (✉)
Division of Biostatistics, Yale University, New Haven, CT 06520, USA
E-mail: shuanggema@yale.edu

M.R. Kosorok
Department of Biostatistics,
University of North Carolina, Chapel Hill, NC 27599, USA

event, but observations are limited to indicators of whether or not the event has occurred at the time the sample is collected. Early examples of current status data arose in demographic applications, with a common version occurring in studies of the distribution of the age at weaning in various settings (Diamond et al. 1986; Grummer-Strawn, 1993). Here, the event of interest is the age of a child at weaning, but the only available information is the weaning status at the observation. Inaccuracy and bias for the measurement of age at weaning, even when weaning occurs before the observation, led to the sole current status on age of weaning at age of observation for the purpose of understanding the distribution of age at weaning. Other examples of current status data arise in carcinogenicity testing when a tumor under investigation is occult (Gart et al., 1986); in the study of infectious diseases, particularly when infection is an unobservable event, as discussed in Jewell and Shiboski (1990) and Shiboski (1998); and in the study of non-fatal human disease, as in Keiding et al. (1996). For a detailed discussion of examples of current status data, see Ma (2004).

Denote Y as the time to event of interest and T as the censoring time. Let Y have a distribution function F. Parametric forms of F may be useful in some situations. The focus of the current research is the study of nonparametric and semiparametric modeling of F because of the robustness and the flexibility. Over the past decade, the asymptotic properties of estimators for nonparametric models with current status data have drawn considerable attention from statistical researchers (Groeneboom and Wellner, 1992; Van der Laan, 1995; Van der Laan et al. 1994). Previous studies conclude that $\hat{F}$, the M-estimator of F, is L_2 consistent, with convergence rate $n^{1/3}$. For any finite sample size, $\hat{F}$ is a step function with jumps at the observed T only. When there exist real valued covariates Z, semiparametric models are usually employed to model the relationship between Y and Z. Examples include (but are not limited to) the Cox model in Huang (1996), the accelerated failure time model in Shen (2000), and the partly linear transformation model in Ma and Kosorok (2004b). It has been shown that the estimators of the nonparametric parameters are L_2 consistent with convergence rate $n^{1/3}$, while the regression parameters can still be estimated $\sqrt{n}$ consistently and asymptotically normally distributed under mild regularity conditions.

Our current study mainly arises from the following concerns. The estimators of the nonparametric parameters for both the nonparametric models (Groeneboom and Wellner, 1992) and the semiparametric models (Huang, 1996; Ma and Kosorok, 2004b) are step functions that converge only at the $n^{1/3}$ rate with the L_2 norm. As pointed out by Hall and Huang (2001), the estimated step functions cannot be used easily for other estimation or inference purposes. Another concern is that, without making stronger continuity assumptions, we cannot achieve uniform consistency even on a compact set (Ma and Kosorok, 2004b). If no differentiability assumptions are made, special computational algorithms for estimating the nonparametric parameters, for example the greatest-convex-minorant algorithm in Huang (1996), are usually needed. Computationally, the discontinuity of the estimated nonparametric functions may cause larger variances for the M-estimators of the regression parameters in semiparametric functions.

Theoretically, it is reasonable to assume the nonparametric parameters are continuously differentiable in certain situations. One possibility is to assume they belong to certain unknown Sobolev spaces (see Wahba, 1990 for reference). Ideally, the gain of making stronger smoothing assumptions would be sharper convergence rates for the estimators of the nonparametric parameters. For semiparametric models, more computationally stable estimators of the regression parameters are expected. Similar smoothness assumptions have been investigated with right censored and interval censored data with penalized splines by Cai and Betensky (2003) and Gu (2002). A difficulty with these approaches is that they are built based on the assumption that the Sobolev spaces are known in advance, which is practically not possible.

The goal of this paper is to take advantage of the smoothness assumptions of the nonparametric parameters, without having to know the underlying Sobolev spaces in advance. To accomplish this, we introduce a modified penalized approach that can yield M-estimators with adaptive convergence rates for estimating the nonparametric parameters. It is shown for semiparametric models that, although the estimators of the nonparametric parameters converge more slowly than $\sqrt{n}$, we can still achieve $\sqrt{n}$ convergence rates and asymptotic normality for estimators of the regression parameters.

The rest of the paper is organized as follows: In Sect. 2, we introduce the modified penalized M-estimators for nonparametric models with current status data and investigate their asymptotic properties. We use the Cox model as a representative of the semiparametric models and study similar penalized estimators in Sect. 3. The focus of Sect. 3 is the asymptotic behaviors of the estimators of the regression parameters. We discuss computational issues in Sect. 4. For illustration, we apply our approach to a small simulation study and to the California Partner Study analysis in the same section. Concluding remarks are in Sect. 5. Some proofs are provided in the Appendix.

2 Modified penalized M-estimation for nonparametric models

Denote Y as the time to event of interest with unknown distribution function F_0, where $F_0 \in \Im_{s_0} = \left\{F\colon \int_l^u \left(F^{(s_0)}(Y)\right)^2 \mathrm{d}Y < \infty\right\}$, the Sobolev space indexed by the order of derivative s_0 for an unknown integer s_0. Superscripts denote derivatives of smooth functions throughout this paper. Suppose observation of Y is subject to censoring T, which is generated from an unknown distribution G. One observation is then composed of $X = (T, \Delta_{(Y\leq T)})$, where Δ is the status indicator. The following data and model assumptions are made.

2.1 Data and model assumptions

It is assumed that:

A1. The censoring time T and event time Y are independent.
A2. $T \in [\tau_l, \tau_u]$, where $0 < \tau_l < \tau_u < \infty$.
A3. There exists a fixed $M > 0$, such that $1/M < F_0$.
A4. $F_0 \in \Im_{s_0}$ and there exists a known integer $s_{\max}$, such that $1 \leq s_0 \leq s_{\max}$.

Condition A1–A3 are standard model assumptions for the nonparametric models with current status data, as discussed in Huang (1996) and Jewell and Van der Laan (2002). It is worth pointing out that condition A3 is necessary only for the maximum likelihood estimator. Assumptions weaker than A2 and A3 can be made for the least squares estimator. Condition A4 is a fairly weak assumption for smooth functions. Under this assumption, the estimates proposed in section 2.2 are at most $s_{\max}^{\text{th}}$ order differentiable. However, there exist "super" smooth functions, i.e., $F_0 \in \Im_s$ for any integer s (for example, the exponential function defined on a compact set). We will discuss this case later. Other smoothness assumptions, for example F_0 belongs to the functional class of local polynomials, can be made. Sobolev spaces are considered here because of their popularity, successes with data analysis, and appealing computational properties (Wahba, 1990).

2.2 The modified penalized M-estimators

Suppose n i.i.d. observations $X_1 = (T_1, \delta_1 = \delta_{(Y_1 \le T_1)}), \ldots, X_n = (T_n, \delta_n = \delta_{(Y_n \le T_n)})$ are generated from the nonparametric model discussed at the beginning of Sect. 2. Based on the n observations, the conditional log-likelihood function takes the form $\log(l(F)) = \sum_{i=1}^{n}\{\delta_i \log(F(T_i)) + (1-\delta_i)\log(1 - F(T_i))\}$. Inspired by previous research by Wahba and Wendelberger (1980) and Van de Geer (2001), we consider the following modified penalized maximum likelihood estimator (MPMLE) $\hat{F}_{\text{MLE}}$, where

$$\hat{F}_{\text{MLE}} = \text{argmax}_{1 \le s \le s_{\max}} \left\{ \text{argmax}_F \left(\log(l(F)) - \text{pen}^2(F) \right) \right\} \tag{1}$$

or the modified penalized least squares estimator (MPLSE) $\hat{F}_{\text{LSE}}$, where

$$\hat{F}_{LSE} = \text{argmin}_{1 \le s \le s_{\max}} \left\{ \text{argmin}_F \left[\left(\sum_{i=1}^{n} (\delta_i - F(T_i))^2 \right) + \text{pen}^2(F) \right] \right\} \tag{2}$$

Denote $F(t) = \int_0^t f(u)\mathrm{d}u = \int_0^t \exp(a(u))\mathrm{d}u$, when $s > 1$. It is easy to see that $a \in \Im_{s-1}$. The penalty $\text{pen}^2(F)$ in Eqs. (1) and (2) is defined as $\text{pen}_s^2(F) = \lambda_n^2(s)(J_s^2(F) + \lambda_0^2)$, where $\lambda_n(s)$ is a data-driven smoothing parameter, $J_s^2(F) = 1\{s=1\}\int_{\tau_l}^{\tau_r} \left(F^{(1)}/F\right)^2 \mathrm{d}t + 1\{s \ne 1\}\int_{\tau_l}^{\tau_r} \left(a^{(s-1)}(t)\right)^2 \mathrm{d}t$, and λ_0 is a model dependent constant that will be discussed later in the proof of Lemma 2.1. The above penalty is modified from ordinary spline settings by including an extra model dependent constant λ_0^2, which plays an important role in controlling the size of the nonparametric estimators. It is also assumed that

B1. $\lambda_n(s) = O_p(n^{-s/(2s+1)})$.
B2. $J_s(\hat{F}_{MLE,LSE}) = o_p(n^{1/12})$.

Condition B1 is usually assumed for estimations of spline functions. In most smoothing spline settings, it is possible to show $J_s(\hat{F}) = O_p(1)$. Although this result has not been proved for our model, it is believed that Condition B2, which requires the penalty on smoothness to grow at a slow rate with the sample size n, can be achieved under most reasonable circumstances.

2.3 Asymptotic properties

The main asymptotic properties of the MPLSE defined in Eq. (2) can be summarized into the following lemma:

Lemma 2.1 *(Adaptive convergence rate of the MPLSE) Under model assumptions A1-A4 and B1-B2,* $||\hat{F}_{LSE} - F_0||_n^2 \equiv \sum_{i=1}^n \left(\hat{F}_{LSE}(t_i) - F_0(t_i)\right)^2 \sim n^{-2s_0/(2s_0+1)}$.

The following definition and technical result are firstly needed:

Definition 2.1 Let $\mathbb{S}$ be a subset of a metric space. The δ-covering number $N(\delta, \mathbb{S})$ is defined as the number of balls with radius δ needed to cover $\mathbb{S}$. The δ-entropy is defined as $H(\delta, \mathbb{S}) = \log N(\delta, \mathbb{S}) \vee 0$. Denote $I(\delta, \mathbb{S}) = \int_0^\delta H^{1/2}(\delta, \mathbb{S})\mathrm{d}\delta$.

Technical Tool 2.1 For all δ, M and integer s,

$$H(u, \{F: ||F - F_0|| \leq \delta, J_s(F) \leq M\}) \\ \leq s \log\left(\frac{5(\delta + M)}{u}\right) + sA^2(M/u)^{1/s},$$

for $0 < u \leq \delta$ and some constant A not depending on s.

Proof Fix $\delta > 0$. Since λ_0 is a constant and $\lambda_n(s) = n^{-s/(2s+1)}$, then for n large enough, $\lambda_n(s) \leq \delta/\lambda_0$ for all $1 \leq s \leq s_{\max}$. From the results shown in the Technical tool, we can see that

$$H(u, \{F: ||F - F_0||^2 + \mathrm{pen}^2(F) \leq \delta^2\}) \\ \leq \sum_{s=1}^{s_{\max}} sA^2\left(\frac{\delta}{u\lambda_s}\right)^{1/s} + s_{\max}\log\left(\frac{10\delta}{u\lambda_{s_{\max}}}\right).$$

From the definition of I, we have

$$\begin{aligned} &I(\delta, \{F: ||F - F_0||^2 + \mathrm{pen}^2(F) \leq \delta^2\}) \\ &\leq \sum_{s=1}^{s_{\max}} \int_0^\delta 2A\sqrt{s}\left(\frac{\delta}{u\lambda_s}\right)^{1/2s}\mathrm{d}u + \sqrt{s_{\max}}\int_0^\delta \log^{1/2}\left(\frac{10\delta}{u\lambda_{\max}}\right)\mathrm{d}u \\ &\leq 2As_{\max}^{3/2}\delta\left(\frac{1}{\lambda_1}\right)^{1/2} + A_0\sqrt{s_{\max}}\delta\sqrt{\log n} \\ &\leq 2As_{\max}^{3/2}\frac{\sqrt{n}\delta^2}{\lambda_0} + A_0\sqrt{s_{\max}}\delta\sqrt{\log n}, \end{aligned} \tag{3}$$

where $A_0 = \int_0^1 \log^{1/2}(10/u)\,\mathrm{d}u$.

Denote $r_i = \delta_i - F_0(T_i)$, for $i = 1, \ldots, n$. From the boundedness conditions of δ_i and F_0, we know $0 \leq r_i \leq 1$, for all i. So for any $K \geq (\log(2))^{-1/2}$, we have $\max_{1\leq i\leq n}\mathrm{E}e^{r_i^2/K^2} \leq 2$. From Theorem 2.1 of Van de Geer (2001), we can then conclude that there exists a constant c, which is a function of K only, such that for $\sqrt{n}\delta_n^2 \geq cI(\delta, \{F: ||F - F_0||^2 + \mathrm{pen}^2(F) \leq \delta^2\})$, we have for all $\delta \geq \delta_n$,

$$\mathrm{P}(||\hat{F}_{\mathrm{LSE}} - F_0||^2 + \mathrm{pen}^2(\hat{F}_{\mathrm{LSE}}) \geq 2(\mathrm{pen}^2(F_0) + \delta^2) \leq c \times \exp\left(-\frac{n\delta^2}{c^2}\right). \tag{4}$$

So we can see if we take

$$\lambda_0^2 \geq 4c^2 A^2 s_{\max}^2, \text{ and } \delta_n = 2cA_0\sqrt{s_{\max}\log\frac{n}{n}},$$

we can conclude that

$$\mathrm{E}||\hat{F}_{\mathrm{LSE}} - F_0||_n^2 + \lambda_n^2(s)(J_s^2(\hat{F}_{\mathrm{LSE}}) + \lambda_0^2) \leq 2\lambda_n^2(s_0)(J_{s_0}^2(F_0) + \lambda_0^2) + d_n^2 + \frac{d_0}{n}, \tag{5}$$

by combining the results in Eq. (3) and (4) with Theorem 2.1 of Van de Geer (2001). In Eq. (5) $d_n = 2A_0\sqrt{s_{\max}\log(n)/n}$, and d_0 is a constant that does not depend on the data or on n. It is obvious that the right hand side of Eq. (5) is dominated by $\lambda_n^2(s_0)$. So Lemma 2.1 holds. □

As discussed in Wahba (1990), the optimal convergence rate for estimators of smooth functions in $\Im_{s_0}$ is $n^{-s_0/(2s_0+1)}$. So the proposed estimator has adaptive convergence rate with respect to the smoothness assumption. The optimal convergence rate for estimating smooth monotone functions depends on the entropy of the corresponding functional set. Unfortunately, that entropy result is still unknown. For $s > 1$, the Sobolev space indexed by s has much smaller size (measured by entropy) compared with the set composed of monotone functions. So it appears likely that the subset of monotone functions in a Sobolev space should have entropy of the same order as the unconstrained Sobolev space. If this conjecture is true, then the proposed estimator has optimal convergence rate. When $s_0 > 1$, the penalized estimator of F has faster convergence rate than the ordinary MLE, which has convergence rate $n^{1/3}$, as discussed in Huang (1996). There exist functions, such as the exponential function defined on finite intervals, that belong to $\Im_s$ for any s. In this case, for any prespecified $s_{\max}$, $||\hat{F}_{\mathrm{LSE}} - F_0||_n^2 \sim n^{-2s_{\max}/(2s_{\max}+1)}$. So theoretically although we do not have full adaptivity in this case, we can achieve the best possible rate (with respect to the smoothness assumption) based on our knowledge in terms of $s_{\max}$ of the smooth functions. Computationally, as suggested in Wahba and Wendelberger (1980), taking $s_{\max} = 5$ or 6 will yield satisfactory results in most situations.

We now investigate the properties of the MPMLE. As can be seen from the proof of Lemma 2.1, the adaptive convergence rate of the penalized least squares estimator for F_0 depends on the entropy results and certain maximal inequalities. Van de Geer (2000) and Ma and Kosorok (2004a) show that for a great variety of semiparametric models the asymptotic behaviors of least squares estimators and MLE are quite similar, if they can be determined by the entropy results. So it is believed that with minor modifications of the proof of Lemma 2.1, we can also show that $||\hat{F}_{\mathrm{MLE}} - F_0||_2^2 \equiv \int_{\tau_l}^{\tau_u}\left(\hat{F}_{\mathrm{MLE}}(t) - F_0(t)\right)^2 \mathrm{d}t \sim n^{-2s_0/(2s_0+1)}$.

2.4 Tuning parameter selection

There are two unknown tuning parameters $\lambda_n(s)$ and λ_0 in the proposed penalization approach. For our study of asymptotic properties, we require $\lambda_n(s)$ to be only of the correct order. From the proof of Lemma 2.1, we can see that λ_0 depends on the constants c and A. A can be determined by closely investigating the entropy bounds, while c is the universal constant in the maximal inequalities (see Lemmas 2.3 and 8.4 of Van de Geer, 2000). Theoretically, these two constants can be determined exactly. On the other hand, it can be seen that if we fix $s_{\max}$, then any non-zero finite λ_0 results in the same asymptotic convergence rates.

Computationally, data driven tuning parameters are expected to better reflect the variability of the data. We propose the following two-step approach based on two fold cross validation. Let $\tilde{I}_n^{(1)} \subset \{1, \dots, n\}$ be a random half of the indices of the observations, and let $\tilde{I}_n^{(2)}$ be the remaining half. When n is odd, the sizes of these two sets will not be equal, but this will not pose any problems. For a given choice of λ_0 and λ_1, let $\hat{F}^{(j)}_{\lambda_0,\lambda_1}$ denote the penalized estimator of F using the objective function in (1) for the data consisting of the observations with indices in $\tilde{I}_n^{(j)}$ and using tuning parameters λ_0 and $\lambda_n(s) = \lambda_1 n^{-1/(2s+1)}$, for $j = 1, 2$. Now compute

$$\tilde{K}_{\lambda_0,\lambda_1} = \sum_{j=1}^{2} \sum_{i \in \tilde{I}_n^{(j)}} \left\{ \delta_i \log\left[\hat{F}^{(3-j)}_{\lambda_0,\lambda_1}(T_i)\right] + (1-\delta_i)\log\left[1 - \hat{F}^{(3-j)}_{\lambda_0,\lambda_1}(T_i)\right]\right\}.$$

We now minimize $\tilde{K}_{\lambda_0,\lambda_1}$ over λ_0 and λ_1 to obtain the cross-validated tuning parameters $\tilde{\lambda}_0$ and $\tilde{\lambda}_1$. Because we only need to achieve the correct scale and not the exact minimizers, we need only vary λ_0 and λ_1 over integral powers of 2. After obtaining $\tilde{\lambda}_0$ and $\tilde{\lambda}_1$ in this manner, we estimate F using the full data with tuning parameters $\lambda_0 = \tilde{\lambda}_0$ and $\lambda_n(s) = \tilde{\lambda}_1 n^{-1/(2s+1)}$.

The above idea can be similarly applied to the least-squares setting and also generalized from two fold to V-fold cross validation. Theoretical studies of V-fold cross validation in other contexts can be found in Smyth (2001). A two-dimensional V-fold cross validation approach somewhat similar to the one we propose has been considered in Gui and Li (2004). For a great variety of spline models, it has been proved that data driven tuning parameters obtained through cross validation can achieve optimal convergence rates. We anticipate that our proposed tuning parameter selection procedure will also yield optimal adaptive rates, but a theoretical verification of this is beyond the scope of the present paper.

For computational simplicity, we set $\lambda_n(s) = n^{-s/(2s+1)}$ and $\lambda_0 = 1$ for the empirical studies in Sects. 4 and 5. Our own limited numerical studies show that fixing the tuning parameters in this manner usually yields satisfactory results.

3 Semiparametric models with current status data

The adaptive convergence rate achieved by using the modified penalized M-estimators for nonparametric models can be naturally extended to semiparametric models. In this section, we use the Cox model (Cox, 1972) with current status data as an example. Most of the results and corresponding proofs do not depend on the special format of the model.

Again let Y and T denote the event time of interest and censoring time, respectively. Moreover, we assume for each subject, a d-dimensional covariate Z is observed. Then a data observation consists of $X = (T, \Delta, Z) \in \mathbb{R}^+ \times \{0, 1\} \times \mathbb{R}^d$.

The density of X is proportional to

$$p_{\beta,\Lambda}(x) = \left(1 - \exp(-e^{\beta' z}\Lambda(t))\right)^{\delta} \left(\exp(-e^{\beta' z}\Lambda(t))\right)^{1-\delta}, \tag{6}$$

where Λ denotes the unknown cumulative baseline hazard function and β is the unknown d-dimensional regression parameter. Following the same argument as for the nonparametric model, we assume the smooth function $\Lambda \in \Im_s$ with unknown s. The main interest here lies in estimation of (β, Λ) and inference for β.

Model Eq. (6) is studied under the following model assumptions:

A1′. The censoring time T and event time Y are conditionally independent given Z.
A2′. $T \in [\tau_l, \tau_u]$, where $0 < \tau_l < \tau_u < \infty$.
A3′. There exists a fixed $M > 0$, such that $1/M < \Lambda_0 < M$.
A4′. $\Lambda_0 \in \Im_{s_0}$, the Sobolev space indexed by the unknown order of derivative s_0, and there exists a known integer $s_{\max}$, such that $1 \le s_0 \le s_{\max}$.
A5. $\beta_0 \in \mathbb{B}_1$ and $Z \in \mathbb{B}_2$, where $\mathbb{B}_1, \mathbb{B}_2$ are known compact sets of $\mathbb{R}^d$.

Assumptions $A1' - A4'$ are parallel to assumptions $A1 - A4$ for the nonparametric model shown in Sect. 2. Assumption $A5$ is the compactness assumption, which will be used for identifiability and entropy control purposes (see Huang, 1996; Ma and Kosorok, 2004b for reference). Following the same scheme as in Sect. 2, we consider the following two M-estimators: the MPMLE defined by

$$(\hat{\beta}_{\text{MLE}}, \hat{\Lambda}_{\text{MLE}}) = \operatorname{argmax}_{1\le s\le s_{\max}} \left\{ \operatorname{argmax}_{\beta,\Lambda} \left(\sum_{i=1}^{n} \log(p_{\beta,\Lambda}(x_i)) - \text{pen}^2(\Lambda) \right) \right\}, \tag{7}$$

and the MPLSE defined by

$$(\hat{\beta}_{\text{LSE}}, \hat{\Lambda}_{\text{LSE}}) = \operatorname{argmin}_{1\le s\le s_{\max}} \left\{ \operatorname{argmin}_{\beta,\Lambda} \left(\sum_{i=1}^{n} \left(1 - \delta_i - \exp(-e^{\beta' z_i}\Lambda(t_i))\right)^2 + \text{pen}^2(\Lambda) \right) \right\}. \tag{8}$$

3.1 Adaptive convergence rate of the semiparametric M-estimators

Semiparametric M-estimators share the same adaptive convergence rate properties as those for the nonparametric model. This is shown below in Lemma 3.1.

Lemma 3.1 *(Adaptive convergence rate of the semiparametric M-estimators) Under assumptions $A1' - A4'$, A5, B1 and*

B2′. $J_s(\hat{\Lambda}_{\text{MLE,LSE}}) = o_p(n^{1/12})$,

it can be proved that $||\hat{\Lambda}_{LSE} - \Lambda_0||_n^2 \equiv \sum_{i=1}^{n} \left(\hat{\Lambda}_{LSE}(t_i) - \Lambda_0(t_i)\right)^2 = O_p(n^{-2s_0/(2s_0+1)})$, and $||\hat{\Lambda}_{MLE} - \Lambda_0||_2^2 \equiv \int_{\tau_l}^{\tau_u} \left(\hat{\Lambda}_{MLE}(t) - \Lambda_0(t)\right)^2 dt = O_p \times (n^{-2s_0/(2s_0+1)})$.

Under the compactness assumption $A5$, it can be shown that the same entropy result as in the proof of Lemma 2.1 holds (see Lemma 3.1 of Huang, 1996). So the adaptive convergence rate for semiparametric models can be proved following the same reasonings as for Lemma 2.1. The details are omitted here.

3.2 Asymptotic behaviors of the estimators of β

A necessary (but not sufficient) condition for the $\sqrt{n}$ consistency and asymptotic normality result for the estimators of the regression parameter is that the variance of the limiting distribution is nonsingular and finite. The variances I_1 and I_2 of the limiting distributions are calculated in the Appendix A. It is assumed that

C1. $0 < \det(I_1) < \infty$,
C1$'$. $0 < \det(I_2) < \infty$,

where det is the determinant of a square matrix.

Lemma 3.2 *($\sqrt{n}$ consistency and asymptotic normality) Under model assumptions $A1' - A4'$, A5, $B1' - B2'$ and C1 (C1$'$), it can be shown that*

$$\sqrt{n}(\hat{\beta}_{MLE} - \beta_0) \to_d N(0, I_1), \ \text{and} \ \sqrt{n}(\hat{\beta}_{LSE} - \beta_0) \to_d N(0, I_2).$$

Moreover, $\hat{\beta}_{MLE}$ is efficient in the sense that any regular estimator has asymptotic variance no less than that of $\hat{\beta}_{MLE}$.

We now consider inference for $\hat{\beta}_{\text{LSE,MLE}}$. Standard methods of semiparametric inferences that depend on the likelihood, such as Huang (1996) and Murphy and Van der Vaart (2000), are not applicable to penalized M-estimators, since the penalty terms are not asymptotically of the order $o(n^{-1})$, as required by the likelihood based approaches. However, the following weighted bootstrap with positive random weights is still valid.

Let $w_1, \ldots, w_n$ be n i.i.d. positive random weights independent of the data. Denote $(\hat{\beta}^*_{\text{MLE}}, \hat{\Lambda}^*_{\text{MLE}})$ as the weighted penalized MLE, i.e.,

$$\begin{aligned}&(\hat{\beta}^*_{\text{MLE}}, \hat{\Lambda}^*_{\text{MLE}})\\&= \text{argmax}_{s \le s_{max}} \left\{ \text{argmax}_{\Lambda \in \Im_s, \beta \in \mathbb{B}_1} \left(\sum_{i=1}^{n} w_i \log(p_{\beta,\Lambda}(x_i)) - \text{pen}^2(\Lambda) \right) \right\}.\end{aligned}$$

We can define $(\hat{\beta}^*_{LSE}, \hat{\Lambda}^*_{LSE})$ in a similar manner.

Lemma 3.3 *(Validity of the weighted bootstrap) Assume that the positive random weights $w_1, \ldots, w_n$ satisfy $E(W) = 1$ and $var(W) = v_0$, where $0 < v_0 < \infty$ is a known constant. We also assume there exists a constant C such that $0 < W < C < \infty$. Then*

$$\left(\sqrt{n/v_0}(\hat{\beta}^*_{\text{MLE,LSE}} - \hat{\beta}_{\text{MLE,LSE}}) \middle| X \right) \to_d \sqrt{n}(\hat{\beta}_{\text{MLE,LSE}} - \beta_0).$$

So the weighted bootstrap provides a valid inference tool.

The validity of the weighted bootstrap depends on the asymptotic behaviors of the weighted penalized M-estimators, which can be verified by applying the techniques used in the proofs of examples 1 and 2 of Ma and Kosorok (2004a). The details are omitted here. The weighted bootstrap with parametric models and semiparametric models with all parameters estimable at the $\sqrt{n}$ rate has been investigated in Barbe and Bertail (1995).

Although we study only the Cox model in detail here, it can be seen that the same technique can be applied to the study of other semiparametric models with current status data, for example the accelerated failure time model (Shen, 2000) and the proportional odds model (Rossini and Tsiatis, 1996). It is expected that under similar model assumptions, the adaptive convergence rate of the nonparametric parameter estimation, $\sqrt{n}$ consistency and asymptotic normality of the estimators of the regression parameters, and validity of inference based on the weighted bootstrap will all follow.

4 Numerical studies

We discuss the computational algorithm and employ a small simulation study with the Cox model discussed in Sect. 3. Similar numerical properties hold for the nonparametric model discussed in Sect. 2. The California Partner Study, where the Cox model is assumed, is analyzed with the proposed penalized approach.

4.1 Computational algorithm

As suggested by Xiang and Wahba (1997), for any fixed s, the maximization over the Sobolev space $\Im_s$ can be achieved by a sieve approach, which states that the sieve estimate, with the number of basis functions growing at least at the rate $n^{1/5}$, can achieve the same asymptotic efficiency as the full space. The K-mean clustering technique is used to select the proper knot positions. Denote k as the number of knots and $\tau_1, \ldots, \tau_k$ as the k data-driven knots. Computationally, we propose $k = \max(20, n^{1/5})$. For $s = 1$, it is assumed that $\Lambda(t) = \mathrm{e}^{a_0} + \mathrm{e}^{a_1} t + \sum_{m=1}^{k} \mathrm{e}^{b_m}(t - \tau_m)_+$, where $(t-\tau_m)_+ = \max(0, t-\tau_m)$ and a_0, a_1 and b_m are unknown coefficients. For $s > 1$, we assume that $f = \sum_{m=1}^{k} c_m B_m(\tau_m)$, where $B_m(\tau_m)$ are the B-spline basis functions, which can be easily generated by R ($http://www.r-project.org$) or S-Plus ($http://www.insightful.com$) and c_ms are unknown coefficients. Since smoothing splines are used here, the same computational concerns as discussed in Wahba (1990) are applicable.

The Newton–Raphson algorithm is applied to maximize over β and the unknown coefficients of Λ. Only simple computations are needed. As shown in the proof of Lemma 3.2, we require our estimators to only "nearly" maximize (minimize) the objective functions. So the Newton–Raphson iterations can be stopped after a finite number of steps.

Besides the desirable asymptotic properties, the proposed approach provides a computationally unified way of estimating semiparametric models with current status data. The Newton–Raphson based algorithm can be applied to various models with only minor modifications.

4.2 Simulation study

We conduct a simulation study to evaluate the performance of the new procedure. We compare our estimator with the MLE studied in Huang (1996). Only the MPMLE is considered here. For simplicity, we consider only a one dimensional covariate. The event times are generated from model Eq. (6), with $\beta_0 = 1$, $Z \sim \text{Unif}[-1.3, 1.3]$ and $\Lambda_0(t) = \int_0^t \exp(u^2/32 - 1)\mathrm{d}u$. The censoring times are assumed to be exponentially distributed, independent of the event times. Our observations are limited to the time interval [0.2, 5]. We simulate 200 realizations for sample sizes equal to 200 and 400. We take $\lambda_0 = 1$.

Figures 1 and 2 show the plot of estimators of β and Λ for sample size equal to 400, based on 200 realizations. The histogram of $\hat{\beta}_{\text{MLE}}$ shows clearly that the

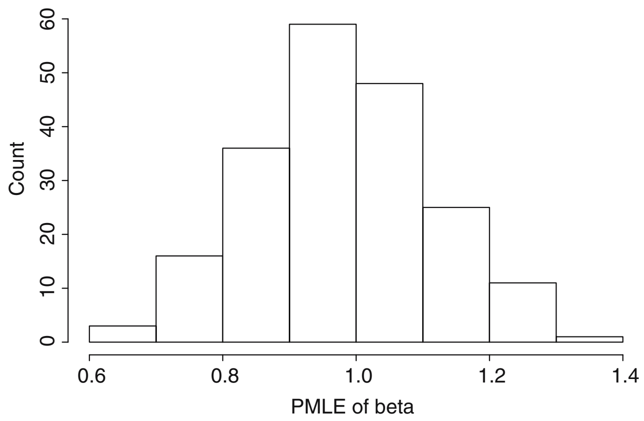

Fig. 1 The histogram of MPMLE $\hat{\beta}_{\text{MLE}}$. Sample size is equal to 400, with 200 realizations

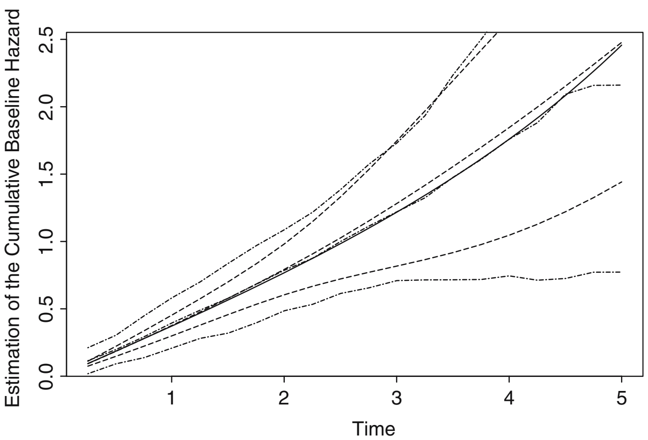

Fig. 2 Simulation study with sample size equal to 400, based on 200 realizations. The plot of the MPMLE and MLE of Λ versus time. The *solid line* is the true cumulative baseline hazard. The *dashed lines* are the MPMLE and corresponding point-wise 95% confidence intervals. The *dot-dashed lines* are the MLE and corresponding point-wise 95% confidence intervals

marginal distribution appears to be Gaussian, as expected. The plot of $\hat{\Lambda}_{\text{MLE}}$ versus time and corresponding point-wise 95% confidence intervals shows satisfying coverage. It can also be seen that the MPMLE has tighter confidence intervals for estimating Λ.

Inference based on the weighted bootstrap is applied. Weights are generated randomly from the exponential distribution with rate equal to 1. The empirical 95% confidence intervals have coverage ratios 0.96 and 0.955 for sample size equal to 200 and 400, respectively, based on 200 bootstraps for each data set. Simulation studies under other data settings show similar satisfactory results.

4.3 Data analysis

Consider the California Partner Study (CPS) of HIV infection (Jewell and Shiboski, 1990). The most straightforward partner study occurs when HIV infection data is collected on both partners in a long-term sexual relationship. Suppose Y denotes the time from infection of the infected case to the infection of the susceptible partner, and that the partnership is evaluated at a single time T after infection of the infected case. Then the infection status of the susceptible partner provides current status data on Y at time T. A schematic representation is available in *Encyclopedia of Biostatistics*.

The partial HIV partner dataset we analyze consists of 302 observations of partners with the male partners as the index cases. The following data analysis is carried out with 295 complete records only. The follow-up time for the 295 partners ranges from 0.08 to 14.9 years. Fifty-five partners developed HIV when monitored. The covariate effect of interest is the average sexual contact rate. A previous analysis in Jewell and Shiboski (1990) suggests log transformation of the covariate. The Cox model has been previously assumed for this study (Jewell and Van der Laan, 2002; Ma, 2004).

We take $\lambda_0 = 1$ and the number of knots $k = 20$. We apply the modified penalized maximum likelihood approach with $s_{\max} = 6$. The MPMLE gives $\hat{\beta} = 0.231$. Using the weighted bootstrap with random weights generated from the exponential distribution with rate 1, the estimated standard error of the estimator is 0.077. These values are close to the MLE values, which are 0.203 and 0.085, respectively. As happened in the simulation study, we obtain a tighter confidence interval for the estimation of β.

5 Concluding remarks

We have proposed a penalized approach that has adaptive convergence rate for estimating the nonparametric parameters for nonparametric and semiparametric models with current status data. The adaptive convergence rate for estimation of functions in Sobolev spaces, together with the efficiency for estimation of the regression parameter, make this penalized approach a valuable alternative to traditional MLE. The adaptive convergence rate achieved by using the modified penalized M-estimators marks significant progress since Gu (2002). The proposed approach also provides a computationally unified way of analyzing semiparametric models with current status data. Computational issues like those discussed in Cai and Betensky (2003) will be investigated in the future.

While our theoretical results are valid for $\lambda_0 = 1$ and $\lambda_n(s) = n^{-1/(2s+1)}$, some improvement for finite samples may be possible by using the two-step, V-fold cross validation procedure proposed in Sect. 2.4. Future studies will be needed to validate this procedure both empirically and theoretically. Based on previous studies of V-fold cross validation in similar penalized contexts Gu, 2002; Wahba, 1990, we expect the proposed procedure to perform satisfactorily.

Our approach and proofs can be applied to right censored and other interval censored data, especially case II interval censored data, with only minor modifications. General transformation models with current status data, right censored data, and interval censored data can also be studied with this penalized approach.

Appendix A

Information calculation

We give the results for the variances of the limiting distributions below. Detailed calculations can be found in Ma and Kosorok (2004a).

For the case of the penalized MLE, we denote

$$Q_{\beta,\Lambda} = \mathrm{e}^{\beta' Z}\left[\delta\frac{\exp(-\mathrm{e}^{\beta' Z}\Lambda)}{1-\exp(-\mathrm{e}^{\beta' Z}\Lambda)} - (1-\delta)\right] \quad \text{and} \quad a_{\beta,\Lambda} = \Lambda\frac{E(ZQ^2_{\beta,\Lambda}|Y)}{E(Q^2_{\beta,\Lambda}|Y)}.$$

Then $I_1 = \left\{\mathrm{P}[(z\Lambda_0 - a_{\beta_0,\Lambda_0})Q_{\beta_0,\Lambda_0}]^{\otimes 2}\right\}^{-1}$.

For the penalized least square estimator, we first make the following notations:

$$\begin{aligned}
m_1 &= 2z\mathrm{e}^{\beta' z}\Lambda\exp(-\mathrm{e}^{\beta' z}\Lambda)(1-\delta-\exp(-\mathrm{e}^{\beta' z}\Lambda)),\\
m_2[a] &= 2\mathrm{e}^{\beta' z}\exp(-\mathrm{e}^{\beta' z}\Lambda)(1-\delta-\exp(-\mathrm{e}^{\beta' z}\Lambda))a,\\
L(\beta,\Lambda) &\equiv 2ze^{\beta' z}\exp(-\mathrm{e}^{\beta' z}\Lambda)\\
&\qquad\times\left((1-\Lambda\mathrm{e}^{\beta' z})(1-\delta-\exp(-\mathrm{e}^{\beta' z}\Lambda)) + \Lambda\mathrm{e}^{\beta' z}\exp(-\mathrm{e}^{\beta' z}\Lambda)\right),\\
N(\beta,\Lambda) &\equiv -2\mathrm{e}^{2\beta' z}\exp(-\mathrm{e}^{\beta' z}\Lambda)\left[1-\delta-2\exp(-\mathrm{e}^{\beta' z}\Lambda)\right], \quad \text{and}\\
A^* &= E(L(\beta,\Lambda)|Y)/E(N(\beta,\Lambda)|Y).
\end{aligned}$$

We also set

$$\begin{aligned}
m_{11} = 2z^2\Lambda\mathrm{e}^{\beta' z}\exp(-\mathrm{e}^{\beta' z}\Lambda)&\left[(1-\Lambda\mathrm{e}^{\beta' z})(1-\delta-\exp(-\mathrm{e}^{\beta' z}\Lambda))\right.\\
&\left.+\Lambda\mathrm{e}^{\beta' z}\exp(-\mathrm{e}^{\beta' z}\Lambda)\right]
\end{aligned}$$

and $m_{21}[a] = L(\beta,\Lambda)a$. Then $I_2 = \{\mathrm{P}(m_{11} - m_{21}[A^*])\}^{-1}\mathrm{P}[m_1 - m_2[A^*]]^2$ $\{\mathrm{P}(m_{11} - m_{21}[A^*])\}^{-1}$.

Table 1 Comparison of MPMLE with MLE: relative bias and empirical standard errors for β_0, $\Lambda(1.25)$, $\Lambda(2.50)$ and $\Lambda(3.75)$ at n=200 and n= 400

	$n=200$				$n=400$			
	β	$\Lambda(1.25)$	$\Lambda(2.50)$	$\Lambda(3.75)$	β	$\Lambda(1.25)$	$\Lambda(2.50)$	$\Lambda(3.75)$
MPMLE								
bias	−0.025	0.021	0.064	−0.017	0.008	0.008	0.043	0.086
S.E.	0.202	0.074	0.196	0.456	0.137	0.050	0.156	0.368
MLE								
bias	0.025	0.042	0.081	0.183	0.008	0.021	0.016	−0.0 04
S.E.	0.212	0.140	0.310	0.655	0.141	0.105	0.193	0.446

S·E: standard error

Appendix B

Proof of Lemma 3.2

The $\sqrt{n}$ consistency and asymptotic normality of $\hat{\beta}_{\text{LSE}}$ and $\hat{\beta}_{\text{MLE}}$ can be proved with the general theorem in Ma and Kosorok (2004a). For simplicity of notation, we consider $\hat{\beta}_{\text{LSE}}$ only. The proof for $\hat{\beta}_{\text{MLE}}$ can be obtained with minor modifications.

Under the assumption B1–B2, we have for any $1 \leq s \leq s_{\max}$, $\lambda_n^2(s)(J_s^2(\hat{\Lambda}_n^1) + \lambda_0^2) = o_p(n^{-1/2})$. So the MPLSE "nearly" minimizes the objective function.

We now check the conditions of Corollary 1 of Ma and Kosorok (2004a). Condition B1, which requires consistency of $\hat{\beta}_{\text{LSE}}$ and the convergence rate of $\hat{\Lambda}_{\text{LSE}}$ to be n^{c_1}, is satisfied from Lemma 2.1 with $c_1 = s_0/(2s_0 + 1)$. Condition B2, which requires finite variance for the limiting distribution, is satisfied by Assumption C1$'$. The smoothness of the model condition B4 can be verified via Taylor expansion techniques for functionals. Since all third derivatives of the objective functions are bounded in a neighborhood of the unknown true value, the smoothness of the model requirement holds with $c_2 = 2$ (in condition B4). For a detailed discussion, see remark 5 of Ma and Kosorok (2004a).

So we need to check only condition B3, the stochastic equicontinuity condition. Denote the functional sets consisting of $m_1(\beta, \Lambda)$ and $m_2[A^*](\beta, \Lambda)$ (for $(\beta, \Lambda) \in \mathbb{S}_1$ defined below) as $\mathbb{M}_1$ and $\mathbb{M}_2$, respectively. Consider the following functional set composed of the LSE estimators

$$\mathbb{S}_1 = \{(\beta, \Lambda) : |\beta - \beta_0| \leq p_1 n^{-s_0/(2s_0+1)},$$

$$||\Lambda - \Lambda_0||_n \leq p_2 n^{-s_0/(2s_0+1)} \text{ and } J_s(\Lambda) = o_p(n^{1/12})\},$$

for two fixed constants $p_1 > 0$ and $p_2 > 0$. Since β belongs to a compact subset of $\mathbb{R}^d$, it can be proved that $\mathbb{S}_1$ has the same entropy result as in Eq. (3). Since $m_1(\beta, \Lambda)$ and $m_2[A^*](\beta, \Lambda)$ defined in the information calculation are differentiable functions of β and Λ, we can conclude that $\mathbb{M}_1$ and $\mathbb{M}_2$ have the same entropy results as $\mathbb{S}_1$, following the same argument as in Lemma 3.1 of Huang (1996). So the stochastic equicontinuity can be proved by applying Theorem 2.14.1 of Van der Vaart and Wellner (1996).

Acknowledgements Both authors were supported in part by grant CA75412 from the National Cancer Institute, USA. We are grateful to the referees, whose insightful comments help to improve an early draft of this article greatly.

References

Barbe, P., Bertail, P. (1995). The weighted bootstrap. New York: Springer

Cai, T., Betensky, R.A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics 59*, 570–579.

Cox, D.R. (1972) Regression models and lifetables (with discussion). *Journal of the Royal Statistical Society, Series B 34*, 187–22.

Diamond, I. D., McDonald, J. W., Shah, I. H. (1986). Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan. *Demography 23*, 607–620.

Gart, J.J., Krewski, D., Lee, P.N., Tarone, R.E., Wahrendorf, J. (1986). *Statistical methods in cancer research, volume III, the design and analysis of longterm Animal experiments*. IARC Scientific Publications N0. 79. Lyon: International Agency for Research on Cancer.

Groeneboom, P., Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Basel: Birkhauser

Grummer-Strawn, L. M. (1993). Regression analysis of current status status data: An application to breast feeding. *Journal of the American Statistical Association 88*, 758–765.

Gu, C. (2002). *Smoothing spline ANOVA models.* Newyork: Springer

Gui, J., Li, K. (2004). Penalized cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Center for Bioinformatics and Molecular Biostatistics*.

Hall, P., Huang, L.S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics 29*, 624–64.

Huang, J. (1996). Efficient estimation for the proportional hazard model with interval censoring. *The Annals of Statistics 24*, 540–568.

Jewell, N. P., Shiboski, S.C. (1990). Statistical analysis of HIV infectivity based on partner studies. *Biometrics 46*, 1133–1150.

Jewell, N.P., Van der Laan, M.J. (2002). Current status data: review, recent developments and open problems. Technical Report: Department of Biostatistics, UC-Berkeley.

Keiding, N., Begtrup, K., Scheike, T.H., Hasibeder, G. (1996). Estimation from current status data in continuous time. *Lifetime Data Analysis 2*, 119–129.

Ma, S. (2004). Penalized M-estimation for partly linear transformation models with current status data. *PhD Thesis, Department of Statistics, UW-Madison*.

Ma, S., Kosorok, M.R. (2004a). Robust semiparametric M-estimation and the weighted bootstrap. *Journal of Multivariate Analysis* (in press).

Ma, S., Kosorok, M.R. (2004b). Penalized log-likelihood estimation for partly linear transformation models with current status data. *Annals of Statistics* (in press).

Murphy, S.A., Van der Vaart, A.W. (2000). On profile likelihood. *Journal of the American Statistical Association 95*, 449–485.

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 93*, 713–721.

Shen, X. (2000). Linear regression with current status data. *Journal of the American Statistical Association 95*, 842–852.

Shiboski, S.C. (1998). Generalized additive models for current status data. *Lifetime Data Analysis 4*, 29–50.

Smyth, P. (2001). Model selection of probabilistic clustering using cross-validated likelihood. *Statistics and Computing* **10**, 63–72.

Van de Geer, S. (2000). *Empirical processes in M-estimation*. Cambridge Series in Statistical and Probabilistic Mathematics.

Van de Geer, S. (2001). Least square estimation with complexity penalties. *Mathematical Methods of statistics* **10**, 355–374.

Van der Laan, M.J. (1995). *Locally efficient estimation with current status data and high dimensional covariates.* Group in Biostatistics Technical Report *55*. University of California, Berkeley.

Van der Laan, M.J., Bickel, P., Jewell, N. P. (1994). Singly and doubly censored current status data: Estimation, asymptotics and regression. *Group in Biostatistics Technical Report 50*, University of California, Berkeley.

Van der Vaart, A.W., Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer

Wahba, G. (1990) *Spline models for observational data.* SIAM. CBMS-NSF Regional Conference Series in Applied Mathematics.

Wahba, G., Wendelberger, J. (1980). Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross Validation *Monthly Weather Review 108*, 1122–1143.

Xiang, D., Wahba, G. (1997) Approximate smoothing spline methods for large data set in the binary case. *Technical Report, Department of Statistics, UW-Madison*.