Lifetime Data Anal (2009) 15:256–277
DOI 10.1007/s10985-008-9110-4

Nonparametric modeling of the gap time in recurrent event data

Pang Du

Received: 20 May 2008 / Accepted: 15 December 2008 / Published online: 3 January 2009

Abstract Recurrent event data arise in many biomedical and engineering studies when failure events can occur repeatedly over time for each study subject. In this article, we are interested in nonparametric estimation of the hazard function for gap time. A penalized likelihood model is proposed to estimate the hazard as a function of both gap time and covariate. Method for smoothing parameter selection is developed from subject-wise cross-validation. Confidence intervals for the hazard function are derived using the Bayes model of the penalized likelihood. An eigenvalue analysis establishes the asymptotic convergence rates of the relevant estimates. Empirical studies are performed to evaluate various aspects of the method. The proposed technique is demonstrated through an application to the well-known bladder tumor cancer data.

Keywords Penalized likelihood · Recurrent event · Gap time hazard function · Asymptotic convergence rate · Bayesian confidence interval

1 Introduction

Recurrent event data arise in many biomedical and engineering studies when failure events can occur repeatedly over time for each subject. Examples of such recurrent event data include hospitalization of schizophrenia patients, infection occurrences among people receiving transplants, tumor recurrences in bladder cancer patients, repeated failures of a particular type of machines, and warranty claims for manufactured products. The development of statistical methods for analyzing such data is therefore of considerable importance. Many existing procedures for recurrent event data model survival and hazard as functions on the domain of calendar time, that is,

P. Du (✉)
Department of Statistics, Virginia Tech, Blacksburg, VA 24061, USA
e-mail: pangdu@vt.edu

time since the start of follow-up. One can find a rich literature review on this topic in Kalbfleisch and Prentice (2002, Chapter 9). Also common in applications is the investigators' interest in the gap time between successive recurrences. See, e.g., Gail et al. (1980). When the covariate is absent, various nonparametric models for the gap time distribution have, among others, been developed by Lin et al. (1999), Wang and Chang (1999) and Peña et al. (2001). In the presence of covariate, semiparametric models have been studied. For example, Strawderman (2005) proposed an accelerated gap time model where the role of the covariate is to accelerate or decelerate a baseline gap time. What is lacking in the literature is a nonparametric model for gap time hazard function in the presence of covariate, which will be explored here.

When modeling hazard function on the calendar time domain, the conventional counting processes counting all the past recurrences have been widely used for asymptotic theory derivation. When the focus switches to hazard function of gap time, it is more convenient to consider processes that restart the time clock every time a recurrence occurs. For example, a gap time counting process counts only the immediate event after the last recurrence; see, e.g., Peña et al. (2001) and Strawderman (2005). The proposed method involves only the gap time processes. Hence, the notation t in this article refers always to the sojourn time, that is, the time since the last recurrence. Being clean of calendar time processes also allows us to include recurrence-specific covariates like the number of past recurrences into the model.

The nonparametric method we consider is based on penalized likelihood. The penalized likelihood for estimating a function of interest η has the form $L(\eta|\text{data}) + \lambda J(\eta)$, where $L(\eta)$, usually the negative log likelihood, measures the goodness-of-fit of η, J, often a quadratic functional, quantifies the smoothness of η, and the smoothing parameter $\lambda(>0)$ controls the tradeoff. For single event data, penalized likelihood has been used in hazard estimation by, e.g., O'Sullivan (1988a,b), Zucker and Karr (1990), Gu (1996), and Joly et al. (1998). For recurrent event data, Rondeau et al. (2007) used penalized likelihood to develop semiparametric joint frailty models of recurrences and death on both the calendar time and the gap time domains.

In modeling recurrent event data, the history of a subject before each recurrence often conveys information for that recurrence. To incorporate the variation between subjects due to their different individual history, a popular approach is to introduce a subject-wise random effect called frailty. Proposed by Vaupel et al. (1979) and popularized by Aalen (1988) and Aalen and Husebye (1991), frailty model with parametric or semiparametric hazard function has been studied extensively in the past decades; see, e.g., Hougaard (2000) for earlier references and Zeng and Lin (2007) for more recent ones. However, in nonparametric estimation by penalized likelihood, the inclusion of random effects can dramatically increase the computational cost and make asymptotic theory elusive; see, e.g., Wang (1998), Karcher and Wang (2001), Gu and Ma (2005a,b) for nonparametric Gaussian and non-Gaussian regression problems with random effects. In this paper, we consider a simpler way to incorporate part of the history by including historical recurrence-wise covariates such as the number of past recurrences, into the analysis. This simplification allows both a faster computation and a thorough investigation of the asymptotic properties of the hazard estimates.

In this paper, we propose a nonparametric model of the gap time hazard function for recurrent event data using penalized likelihood. We first derive a score for smoothing

parameter selection through delete-one-subject cross-validation and properties of the gap-time scale counting processes. Then for inference purpose, point-wise confidence intervals of gap time hazard function are developed from the Bayes model associated with penalized likelihood. The asymptotic convergence rates for our estimates are also proved assuming an eigenvalue analysis with suitable properties. Our simulations investigate several aspects of the proposed method. Combined with the model selection tool developed in Gu (2004) for penalized likelihood estimation, we then apply the proposed method to the well-known bladder tumor cancer data in Byar (1980).

The rest of the article is organized as follows. Section 2 illustrates the penalized likelihood method for estimating gap time hazard function in recurrent event data, with Sect. 2.1 introducing the model, Sect. 2.2 giving estimation details, Sect. 2.3 deriving the cross-validation score for smoothing parameter selection, Sect. 2.4 developing point-wise confidence intervals, and Sect. 2.5 showing the convergence rates of the estimates. Section 3 devotes to numerical studies, where the smoothing parameter selection and the confidence intervals are examined. In Sect. 4, we apply the proposed method to the well-known bladder tumor cancer data in Byar (1980). Remarks in the last section conclude the article.

2 Penalized likelihood method

2.1 Model

Consider n independent subjects. Of each one observes a sequence of events at times

$$0 < S_{i1} < S_{i2} < \cdots < S_{iK_i} \leq C_i,$$

where C_i is the right censoring time independent of S_{ik}'s. The inter-occurrence gap times are denoted by

$$T_{i1} = S_{i1}, \quad T_{i2} = S_{i2} - S_{i1}, \ldots, T_{iK_i} = S_{iK_i} - S_{i,K_i-1}.$$

Assume each gap time T_{ij} depends only on a covariate U_{ij} such that the $T_{ij}|U_{ij}$ are a random sample of the random variable $T|U$ with survival function $S(t,u) = \text{Prob}(T > t|U = u)$. A similar semiparametric gap-time-covariate dependence structure is available in Strawderman (2005). Also note that our covariate, which may change only at recurrences, is specific to each recurrent event rather than each subject. This is necessary for us to incorporate part of a subject's history by including covariates like the number of past recurrences. Of interest is the estimation of the gap time hazard function

$$h(t,u) = -\partial \log S(t,u)/\partial t.$$

Let's first introduce some auxiliary symbols to simplify our future notation. For each i, let

$$\Xi_i = \begin{cases} K_i + 1 & \text{if } S_{iK_i} = C_i \text{ (censored)}, \\ K_i & \text{if } S_{iK_i} < C_i \text{ (uncensored)}. \end{cases}$$

When subject i is censored such that $S_{iK_i} = C_i$, let S_{i,K_i+1} be the unobserved $(K_i + 1)$-st event time for the subject. Then $T_{i,K_i+1} = S_{i,K_i+1} - S_{iK_i}$ would be the $(K_i + 1)$-st gap time. Now, C_{ij}, defined by

$$C_{ij} \stackrel{\text{def}}{=} C_i - S_{i,j-1} = C_i - \sum_{k=1}^{j-1}(S_{ik} - S_{i,k-1}) = C_i - \sum_{k=1}^{j-1} T_{ik},$$

acts like a "right censoring" time for T_{ij}. Indeed, C_{ij} is independent of T_{ij} (but not of $T_{ik}, k \le j-1$). Although the only gap time that is possibly "censored" is T_{i,K_i+1}, the model may now be restated as a right-censored data problem, as follows. For $i = 1, 2, \ldots, n$, we observe

$$X_{ij} = \min(T_{ij}, C_{ij}) \quad \text{and} \quad \delta_{ij} = I\{T_{ij} \le C_{ij}\}, \quad j = 1, \ldots, \Xi_i.$$

A simple but informal approach in Sect. 3 of Aalen and Husebye (1991) expresses the contribution of subject i to the likelihood function as

$$\prod_{j=1}^{K_i} P\{j\text{th recurrence at } T_{ij} | U_{ij}, U_{ik}, X_{ik}, k < j\}$$
$$\times [P\{(K_i+1)\text{st recurrence censored at } C_i | C_i, U_{ik}, X_{ik}, k = 1, \ldots, K_i\}]^{1-\delta_{i,\Xi_i}}, \tag{1}$$

which are really hybrids of probability densities for the uncensored recurrences and a genuine probability for the censored recurrence. Although a more precise measure-theoretic framework is not used in (1), its meaning should be intuitively clear. By our conditional independence assumptions on X_{ij}'s, the contribution of each uncensored recurrence X_{ij} $(= T_{i,j})$, $j = 1, \ldots, K_i$ is simply $f(X_{ij}, U_{ij})$ with $f(t, u)$ being the density function of $T|U$. A censored recurrence X_{i,K_i+1} $(= C_{i,K_i+1})$ gives

$$P\{(K_i+1)\text{-st recurrence time} > C_{i,K_i+1} | C_{i,K_i+1}, U_{ik}, X_{ik}, k = 1, \ldots, K_i\}$$
$$= P\{(K_i+1)\text{-st recurrence time} > X_{i,K_i+1} | U_{iK_i}\} = S(X_{i,K_i+1}, U_{iK_i}).$$

Note that $f(t, u) = h(t, u)S(t, u)$ and $S(t, u) = \exp(-\int_0^t h(s, u)ds)$. Then the contribution of subject i to the likelihood function is

$$\prod_{j=1}^{\Xi_i} h(X_{ij}, U_{ij})^{\delta_{ij}} \exp\left(-\int_0^{X_{ij}} h(t, U_{ij})dt\right),$$

where $U_{i,K_i+1} = U_{i,K_i}$ for the censored gap time X_{i,K_i+1}. Let $N = \sum_{i=1}^n \Xi_i$. Thus, the *penalized likelihood* estimate of $\eta_0(t, u) = \log h(t, u)$ is defined to be the minimizer of

$$-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\left\{\delta_{ij}\eta(X_{ij},U_{ij})-\int_0^{X_{ij}}e^{\eta(t,U_{ij})}dt\right\}+\frac{\lambda}{2}J(\eta), \tag{2}$$

in a Hilbert space $\mathcal{H}$ of functions on the product domain $\mathcal{T}\times\mathcal{U}$ of gap time and covariate. Here the double sum term is the negative log likelihood, J is a roughness penalty functional, and the smoothing parameter $\lambda>0$ controls the tradeoff between the goodness-of-fit and smoothness of the estimate.

Most recurrent event data contain covariate, so η is often a function of gap time and covariate. With a generic covariate u that can be a vector of continuous and/or discrete covariates, a functional ANOVA decomposition of η is

$$\eta(t,u)=\eta_0+\eta_t(t)+\eta_u(u)+\eta_{t,u}(t,u), \tag{3}$$

where η_0 is a constant like the grand mean, η_t is the main effect of time t, η_u is the main effect of covariate u, and $\eta_{t,u}(t,u)$ is the interaction effect between time and covariate. When $\eta_{t,u}=0$, (3) reduces to an additive model for η, or the well-known proportional hazards model. Various side conditions through averaging operators, such as the side conditions $\int_{\mathcal{T}}\eta_t(t)dt=0$, $\int_{\mathcal{U}}\eta_u(u)du=0$, $\int_{\mathcal{T}}\eta_{t,u}(t,u)dt=0$ and $\int_{\mathcal{U}}\eta_{t,u}(t,u)du=0$ on η_t, η_u and $\eta_{t,u}$, are needed to ensure the identifiability of the terms in (3); see Wahba (1990) and Gu (2002).

2.2 Estimation

The minimization of (2) is performed in a Hilbert space $\mathcal{H}\subset\{\eta: J(\eta)<\infty\}$ on the product domain $\mathcal{T}\times\mathcal{U}$ of gap time and covariate, in which J is a square semi-norm. The evaluation functional $[(t,u)]f=f(t,u)$ is assumed to be continuous in $\mathcal{H}$, which is necessary for (2) to be continuous in its argument η. A Hilbert space $\mathcal{H}$ in which the evaluation functional is continuous is called a reproducing kernel Hilbert space (RKHS). It possesses a reproducing kernel $R(\cdot,\cdot)$, a nonnegative definite function with the reproducing property that $\langle R((t,u),\cdot),f(\cdot)\rangle=f(t,u)$ for any $f\in\mathcal{H}$, where $\langle\cdot,\cdot\rangle$ is the inner product in $\mathcal{H}$. The null space $\mathcal{N}_J$ of J in $\mathcal{H}$ is finite dimensional to prevent interpolation, the conceptual equivalent of a delta sum. Then $\mathcal{H}$ can be decomposed into a tensor sum $\mathcal{N}_J\oplus\mathcal{H}_J$, with $\mathcal{H}_J$ possessing a reproducing kernel $R_J(\cdot,\cdot)$. See Aronszajn (1950), Wahba (1990) and Gu (2002) for more details on RKHS.

Example 1 (Cubic Spline) In the absence of covariate, $\mathcal{U}$ decays to a singleton. One only has the gap time domain $\mathcal{T}$, say $[0,1]$. A choice of $J(\eta)$ is $\int_0^1(\eta'')^2dt$, which yields the popular cubic splines. If the inner product in $\mathcal{N}_J$ is $(\int_0^1 f\,dt)(\int_0^1 g\,dt)+(\int_0^1 f'\,dt)(\int_0^1 g'\,dt)$, then $\mathcal{H}_J=\{\eta:\int_0^1\eta dt=\int_0^1\eta' dt=0, J(\eta)<\infty\}$ and the reproducing kernel $R_J(t_1,t_2)=k_2(t_1)k_2(t_2)-k_4(|t_1-t_2|)$, where $k_\nu(t)=B_\nu(t)/\nu!$ are scaled Bernoulli polynomials for $t\in[0,1]$. The null space $\mathcal{N}_J$ has a basis $\{1,k_1(t)\}$, where $k_1(t)=t-0.5$ for $t\in[0,1]$. See Gu (2002, Sect. 2.3.3). □

In the presence of covariate, one needs to estimate a multivariate function η. To incorporate the functional ANOVA structure in (3), we consider the univariate function

$\eta_* \in \mathcal{H}_{\langle * \rangle}$, where $*$ stands for t or u. $\mathcal{H}_{\langle * \rangle}$ is a RKHS with tensor sum decomposition $\mathcal{H}_{\langle * \rangle} = \mathcal{H}_{0\langle * \rangle} \oplus \mathcal{H}_{1\langle * \rangle}$. $\mathcal{H}_{0\langle * \rangle}$ is the finite dimensional "parametric" subspace consisting of parametric functions, and $\mathcal{H}_{1\langle * \rangle}$ is the "nonparametric" subspace consisting of nonparametric smooth functions. This yields a tensor product space

$$\begin{aligned}\mathcal{H} &= \mathcal{H}_{\langle t \rangle} \otimes \mathcal{H}_{\langle u \rangle} \\ &= (\mathcal{H}_{0\langle t \rangle} \otimes \mathcal{H}_{0\langle u \rangle}) \oplus (\mathcal{H}_{1\langle t \rangle} \otimes \mathcal{H}_{0\langle u \rangle}) \oplus (\mathcal{H}_{0\langle t \rangle} \otimes \mathcal{H}_{1\langle u \rangle}) \oplus (\mathcal{H}_{1\langle t \rangle} \otimes \mathcal{H}_{1\langle u \rangle}),\end{aligned} \tag{4}$$

where the components of the tensor sum are all on the product domain $\mathcal{T} \times \mathcal{U}$, and parallel to the components in the functional ANOVA decomposition (3).

Example 2 (tensor product spline) When the covariate is a categorical variable with l levels, the covariate domain becomes $\mathcal{U} = \{1, \ldots, l\}$. Functions on $\mathcal{U}$ are essentially vectors in R^l, so the RKHS $\mathcal{H}_{\langle u \rangle} = R^l$.

For nominal covariate, let $\bar{\eta} = \sum_{x=1}^{l} f(x)/l$. Equipped with the roughness penalty $J_{\langle u \rangle}(\eta) = \sum_{x=1}^{l} [\eta(x) - \bar{\eta}]^2$ and inner product $\langle f, g \rangle = \sum_{x=1}^{l} f(x)g(x)$, the RKHS $\mathcal{H}_{\langle u \rangle}$ decomposes as

$$\mathcal{H}_{\langle u \rangle} = \mathcal{H}_{0\langle u \rangle} \oplus \mathcal{H}_{1\langle u \rangle} = \{\eta : \eta(1) = \cdots = \eta(l)\} \oplus \left\{ \eta : \sum_{x=1}^{l} \eta(x) = 0 \right\}$$

with reproducing kernels $R_{0\langle u \rangle}(u_1, u_2) = 1/l$, $R_{1\langle u \rangle}(u_1, u_2) = I_{[u_1 = u_2]} - 1/l$.

For ordinal covariate, a natural alternative roughness penalty is $J_{\langle u \rangle}(\eta) = \sum_{x=2}^{l} [\eta(x) - \eta(x-1)]^2$ which yields the same decomposition of $\mathcal{H}_{\langle u \rangle}$ but with reproducing kernels $R_0(u_1, u_2) = (I - B)^+_{u_1, u_2}$ and $R_1(u_1, u_2) = B^+_{u_1, u_2}$. Here $M^+_{i,j}$ is the (i, j)-th entry of the Moore–Penrose inverse of a matrix M and

$$B = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}_{l \times l}.$$

On the other hand, the construction in Example 1 gives a decomposition of the RKHS $\mathcal{H}_{\langle t \rangle}$ on the gap time domain

$$\begin{aligned}\mathcal{H}_{\langle t \rangle} &= \left\{ \eta : \int_0^1 (\eta'')^2 dx < \infty \right\} = \mathcal{H}_{0\langle t \rangle} \oplus \mathcal{H}_{1\langle t \rangle} \\ &= \text{span}\{1, k_1(x)\} \oplus \left\{ \eta : \int_0^1 \eta dx = \int_0^1 \eta' dx = 0, \ \int_0^1 (\eta'')^2 dx < \infty \right\},\end{aligned}$$

with reproducing kernels $R_{0\langle t\rangle}(t_1, t_2) = 1 + k_1(t_1)k_1(t_2)$, and $R_{1\langle t\rangle}(t_1, t_2) = k_2(t_1)k_2(t_2) - k_4(|t_1 - t_2|)$. The tensor product of $\mathcal{H}_{\langle t\rangle}$ and $\mathcal{H}_{\langle u\rangle}$ yields four tensor sum terms $\mathcal{H}_{\nu,\mu} = \mathcal{H}_{\nu\langle t\rangle} \otimes \mathcal{H}_{\mu\langle u\rangle}$ on $\mathcal{T} \times \mathcal{U}$, $\nu = 0, 1$ and $\mu = 0, 1$, with reproducing kernels $R_{\nu,\mu}(x_1, x_2) = R_\nu(t_1, t_2)R_\mu(u_1, u_2)$, where $x_i = (t_i, u_i)$. The subspace with $\nu = 0$ and $\mu = 0$ becomes the null space $\mathcal{N}_J$. The other three subspaces form $\mathcal{H}_J$ with the reproducing kernel

$$R_J = \theta_{0,1}R_{0\langle t\rangle,1\langle u\rangle} + \theta_{1,0}R_{1\langle t\rangle,0\langle u\rangle} + \theta_{1,1}R_{1\langle t\rangle,1\langle u\rangle},$$

where $\theta_{\nu,\mu}$ are a set of extra smoothing parameters adjusting the relative weights of the roughness of different components. See Gu (2002, Chapter 2). □

Since $\mathcal{H}$ is usually of infinite dimensions, the minimization of (2) in $\mathcal{H}$ is generally uncomputable. To circumvent the problem, one can perform the minimization in a data-adaptive finite dimensional space $\mathcal{H}_n = \mathcal{N}_J \oplus \text{span}\{R_J((X_{ij}, U_{ij}), \cdot) : j = 1, \ldots, K_i; i = 1, \ldots, n\}$. Under mild conditions, the minimizers in $\mathcal{H}_n$ can be shown to share the same asymptotic convergence rates as the minimizers in $\mathcal{H}$; see Sect. 2.5.

Let's re-index the set $\{(X_{ij}, U_{ij}) : j = 1, \ldots, K_i; i = 1, \ldots, n\}$ as $\{v_k : k = 1, \ldots, N_n\}$ with $N_n = \sum_{i=1}^n K_i$ such that $v_1 = (X_{11}, U_{11})$, $v_2 = (X_{12}, U_{12})$, and so on. Write $\xi_k = R_J(v_k, \cdot)$ and let $\{\phi_\nu\}_{\nu=1}^m$ be a basis of $\mathcal{N}_J$. By definition, a function in $\mathcal{H}_n$ has an expression

$$\eta(t, u) = \sum_{\nu=1}^{m} d_\nu\phi_\nu(t, u) + \sum_{k=1}^{N_n} c_k R_J(v_k, (t, u)) = \boldsymbol{\phi}^T\mathbf{d} + \boldsymbol{\xi}^T\mathbf{c}, \tag{5}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\xi}$ are vectors of functions and $\mathbf{d}$ and $\mathbf{c}$ are vectors of coefficients. Substituting (5) into (2), one calculates the minimizer of (2) in $\mathcal{H}_n$ by minimizing

$$A_\lambda(\mathbf{c}, \mathbf{d}) = -\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\left\{\delta_{ij}(\boldsymbol{\phi}_{ij}^T\mathbf{d} + \boldsymbol{\xi}_{ij}^T\mathbf{c}) - \int_0^{X_{ij}} \exp(\boldsymbol{\phi}(t, U_{ij})^T\mathbf{d} + \boldsymbol{\xi}(t, U_{ij})^T\mathbf{c})dt\right\} + \frac{\lambda}{2}\mathbf{c}^T Q\mathbf{c} \tag{6}$$

with respect to $\mathbf{d}$ and $\mathbf{c}$, where $\boldsymbol{\phi}_{ij}$ is $m \times 1$ with the νth entry $\phi_\nu(X_{ij}, U_{ij})$, $\boldsymbol{\xi}_{ij}$ is $N_n \times 1$ with the kth entry $\xi_k(X_{ij}, U_{ij})$ and Q is $N_n \times N_n$ with the (j, k)th entry $R_J(v_j, v_k)$.

Write $\mu_f(g) = (1/N)\sum_{i=1}^n \sum_{j=1}^{\Xi_i} \int_0^{X_{ij}} g(t, U_{ij})e^{f(t,U_{ij})}dt$, $V_f(g, h) = \mu_f(gh)$, and $e(f) = (1/N)\sum_{i=1}^n \sum_{j=1}^{K_i} f(X_{ij}, U_{ij})$. The minimization of (6) is carried out by two nested loops. For fixed smoothing parameters, the inner loop minimizes (6) through the Newton–Raphson procedure, which updates the coefficients from the current iterate $\tilde{\eta} = \boldsymbol{\phi}^T\tilde{\mathbf{d}} + \boldsymbol{\xi}^T\tilde{\mathbf{c}}$ through

$$\begin{pmatrix} V_{\phi,\phi} & V_{\phi,\xi} \\ V_{\xi,\phi} & V_{\xi,\xi} + \lambda Q \end{pmatrix}\begin{pmatrix} \mathbf{d} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} e_\phi - \mu_\phi + V_{\phi,\eta} \\ e_\xi - \mu_\xi + V_{\xi,\eta} \end{pmatrix}, \tag{7}$$

where $V_{\phi,\phi} = V_{\tilde{\eta}}(\boldsymbol{\phi}, \boldsymbol{\phi}^T)$, $V_{\phi,\xi} = V_{\xi,\phi}^T = V_{\tilde{\eta}}(\boldsymbol{\phi}, \boldsymbol{\xi}^T)$, $V_{\xi,\xi} = V_{\tilde{\eta}}(\boldsymbol{\xi}, \boldsymbol{\xi}^T)$, $\mu_\phi = \mu_{\tilde{\eta}}(\boldsymbol{\phi})$, $\mu_\xi = \mu_{\tilde{\eta}}(\boldsymbol{\xi})$, $e_\xi = e(\boldsymbol{\xi})$, $e_\phi = e(\boldsymbol{\phi})$, $V_{\phi,\eta} = V_{\tilde{\eta}}(\boldsymbol{\phi}, \tilde{\eta})$, and $V_{\xi,\eta} = V_{\tilde{\eta}}(\boldsymbol{\xi}, \tilde{\eta})$. The selection of the smoothing parameters can be done through an outer-loop optimization of a cross-validation score derived in Sect. 2.3.

2.3 Smoothing parameter selection

With varying smoothing parameters λ (and $\theta_{v,\mu}$'s), the minimizers of (2) define an array of possible estimates. To choose an optimal estimate, one must select λ through a data-adaptive approach; see, e.g., Gu (1996) and Joly et al. (1998).

Let η_λ be the minimizer of the penalized likelihood (2). Define the Kullback–Leibler distance between the true log hazard η_0 (not to be confused with the MLE $\eta_\lambda|_{\lambda=0}$) and η_λ to be

$$\begin{aligned}&\mathrm{KL}(\eta_0, \eta_\lambda)\\ &= E\left[\int_{\mathcal{T}} Y(t)\left\{(\eta_0(t,U) - \eta_\lambda(t,U))e^{\eta_0(t,U)} - (e^{\eta_0(t,U)} - e^{\eta_\lambda(t,U)})\right\} dt\right], \quad (8)\end{aligned}$$

where $Y(t) = I_{[X \geq t]}$ is the at-risk process and the expectation is with respect to X and U. Let $\mathrm{SKL}(\eta_0, \eta_\lambda) = \mathrm{KL}(\eta_0, \eta_\lambda) + \mathrm{KL}(\eta_\lambda, \eta_0)$ be the symmetrized Kullback–Leibler distance between η_0 and η_λ as defined in (12). Under general conditions, we will show, in Sect. 2.5, $\mathrm{SKL}(\eta_0, \eta_\lambda) \to 0$ as $N \to \infty$. Estimating (8) through a combination of the gap time counting processes and a delete-one-subject cross-validation, one ends up with the score

$$\begin{aligned}V_\alpha(\lambda) = &-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\left\{\delta_{ij}\eta_\lambda(X_{ij}, U_{ij}) - \int_0^{X_{ij}} e^{\eta_\lambda(t,U_{ij})}dt\right\}\\ &+\alpha\cdot\frac{1}{N}\left\{\mathbf{1}^T(I/N - D)\breve{Q}^T H^{-1}\breve{Q}\mathbf{1} + \mathrm{tr}(D\breve{Q}^T H^{-1}\breve{Q})\right\}, \quad (9)\end{aligned}$$

where $\breve{Q} = (\sum_{j=1}^{K_1}\boldsymbol{\psi}(X_{1j}, U_{1j}), \ldots, \sum_{j=1}^{K_n}\boldsymbol{\psi}(X_{nj}, U_{nj}))$ with $\boldsymbol{\psi} = (\boldsymbol{\xi}^T, \boldsymbol{\phi}^T)^T$, I is the identity matrix and $D = \mathrm{diag}\left(\frac{1}{N-\Xi_1}, \ldots, \frac{1}{N-\Xi_n}\right)$. See Appendix A for detailed derivation. In (9), the constant $\alpha > 1$ is added to prevent occasional severe under-smoothing typically suffered by cross-validation methods. For hazard estimation of single event data, an α around 1.4 was suggested by Gu (2002) to ensure little loss of effectiveness. A similar range is indicated by our empirical studies in Sect. 3.

2.4 Bayesian confidence intervals

Due to the lack of parametric sampling distribution, a rigorously justified interval estimate is a rarity in nonparametric function estimation. An example in point is the Bayesian confidence intervals of Wahba (1983), which are derived from the Bayes

model of smoothing splines. Following this direction, interval estimates of hazard function in single event survival data were developed in, e.g., Joly et al. (1998) and Du and Gu (2006). In this section, we shall derive the Bayesian confidence intervals for the hazard estimate of (2). Empirical coverage properties of the intervals will be investigated in Sect. 3.2.

Consider $\eta = \eta_0 + \eta_1$, where η_0 has a diffuse prior in $\mathcal{N}_J$ and η_1 has a mean 0 Gaussian process prior with the covariance function

$$E[\eta_1(x_1)\eta_1(x_2)] = bR_J(x_1, \mathbf{v}^T)Q^+R_J(\mathbf{v}, x_2),$$

where $x_1, x_2 \in \mathcal{T} \times \mathcal{U}$ and Q^+ is the Moore–Penrose inverse of $Q = R_J(\mathbf{v}, \mathbf{v}^T)$; one may parameterize $\eta_0 = \sum_{\nu=1}^{m} d_\nu \phi_\nu$ with d_ν diffuse and $\eta_1 = \sum_{k=1}^{N_q} c_k \xi_k$ with $\mathbf{c} \sim N(\mathbf{0}, bQ^+)$. Setting $b = (N\lambda)^{-1}$, the minimizer η_λ of (2) in $\mathcal{H}_n$ is seen to be the posterior mode under this prior. The Bayesian confidence intervals are based on a quadratic approximation of the log likelihood at this posterior mode.

Recall the expression $\eta = \boldsymbol{\phi}^T\mathbf{d} + \boldsymbol{\xi}^T\mathbf{c}$ in (5). To simplify the notation, let's write $\boldsymbol{\psi} = (\boldsymbol{\xi}^T, \boldsymbol{\phi}^T)^T$ and $\check{\mathbf{c}} = (\mathbf{d}^T, \mathbf{c}^T)^T$. Through a second order Taylor expansion of the integral terms of (6), its quadratic approximation at $\tilde{\eta} = \eta_\lambda$ is seen to be

$$\frac{1}{2N}(\check{\mathbf{c}} - \check{\mathbf{c}}_\lambda)^T(NH)(\check{\mathbf{c}} - \check{\mathbf{c}}_\lambda) + C, \tag{10}$$

where H is the left-hand side matrix in (7), $\check{\mathbf{c}}_\lambda$ is the coefficient vector of η_λ, and C is a constant. Note that (5) and (6) are the negative log posterior divided by N, so the approximate posterior of $\check{\mathbf{c}} = (\mathbf{d}^T, \mathbf{c}^T)^T$ through (10) is Gaussian with mean $\check{\mathbf{c}}_\lambda$ and covariance H^+/N. It follows that given a point $x = (t, u)$, the approximate posterior mean of $\eta(x)$ is $\eta_\lambda(x) = \boldsymbol{\phi}^T(x)\mathbf{d}_\lambda + \boldsymbol{\xi}^T(x)\mathbf{c}_\lambda = \boldsymbol{\psi}^T(x)\check{\mathbf{c}}_\lambda$ and the approximate posterior variance is $s^2(x) = \boldsymbol{\psi}^T(x)H^+\boldsymbol{\psi}(x)/N$. Then the Bayesian confidence interval at $x = (t, u)$ is $\eta_\lambda(x) \pm z_{\alpha/2}s(x)$.

Note that the Bayesian confidence interval is point-wise. Although Nychka (1988) has shown certain across-the-function coverage property for smoothing splines in Wahba (1983), it is unclear whether this can be extended to our case.

2.5 Asymptotic convergence

In this section we shall look at the asymptotic convergence properties of the penalized likelihood estimates. Define $V(f) = V(f, f)$ with

$$V(f, g) = \int_{\mathcal{U}} m(u) \int_{\mathcal{T}} f(t, u)g(t, u)e^{\eta_0(t,u)}S(t, u)dt, \tag{11}$$

where $m(u)$ is the density function of U. Denote the symmetrized Kullback–Leibler distance between η and η_0 by

$$\mathrm{SKL}(\eta_0, \eta) = \int_{\mathcal{U}} m(u) \int_{\mathcal{T}} (e^{\eta(t,u)} - e^{\eta_0(t,u)})(\eta(t, u) - \eta_0(t, u))S(t, u)dt. \tag{12}$$

The convergence rates for an estimate η will be established in terms of $V(\eta - \eta_0)$ and $\mathrm{SKL}(\eta_0, \eta)$ through an eigenvalue analysis. The following smoothness and regularity conditions are assumed for the main result; see Gu (2002, Chapter 8) for similar conditions for hazard estimation of single event survival data.

Condition 1: V is completely continuous with respect to J.
Condition 2: For ν sufficiently large and some $\beta > 0$, the eigenvalues π_ν of J with respect to V satisfy $\pi_\nu > \beta\nu^r$, where $r > 1$.
Condition 3: Let $\hat{\eta}$ and $\hat{\eta}^*$ be respectively the minimizers of the penalized likelihood functional (23) in $\mathcal{H}$ and $\mathcal{H}_n$. For η in a convex set B_0 around η_0 containing $\hat{\eta}$, $\hat{\eta}^*$ and two other intermediate estimates to be defined below, assume $c_1 e^{\eta_0(t,u)} \le e^{\eta(t,u)} \le c_2 e^{\eta_0(t,u)}$ holds uniformly for some $c_1 > 0$ and $c_2 < \infty$ on $\{(t,u) : S(t,u) > 0\}$.
Condition 4: $\int_{\mathcal{U}} m(u) \int_{\mathcal{T}} \zeta_\nu^2 \zeta_\mu^2 e^{k\eta_0} S dt \le c_3$ for some $c_3 < \infty$, $k = 1, 2$, $\forall \nu, \mu$.
Condition 5: $\sum_\nu \pi_\nu^p \eta_{\nu,0}^2 < \infty$ for some $p \in [1, 2]$.

One can refer to Weinberger (1974) for the exact definition of one quadratic functional being completely continuous with respect to another quadratic functional. Here, Condition 1 is to bound λJ for fixed λ such that the effective model space dimension can be kept finite to achieve noise reduction in estimation. Also, the effective model space dimension should be expandable as data become available. This is achieved by letting $\lambda \to 0$ as $n \to \infty$. Condition 2 essentially dictates how fast λ should approach 0 by restricting the growth rate of the eigenvalues π_ν. The constant r there quantifies the smoothness imposed by $J(\eta) : r = 4$ for the cubic spline of Example 1, and $r = 4 - \delta, \forall \delta > 0$ for the tensor product cubic spline of Example 2. By the mean value theorem, Condition 3 implies the equivalence of $V(\eta - \eta_0)$ and $\mathrm{SKL}(\eta_0, \eta)$ for η in B_0. When η_0 is bounded, Condition 4 essentially requires a uniform bound on the fourth moments of ζ_ν. Condition 5 describes the smoothness of η_0: when $p = 1$, $\sum_\nu \pi_\nu \eta_{\nu,0}^2 = J(\eta_0) < \infty$ is equivalent to say that $\ddot{\eta}_0$ is square integrable; $p = 2$ means that $\eta_0^{(4)}$ is square integrable or η_0 is "supersmooth".

Theorem 1 *Let $\hat{\eta}$ and $\hat{\eta}^*$ be respectively the minimizers of the penalized likelihood functional (23) in $\mathcal{H}$ and $\mathcal{H}_n$. Under Conditions 1–5, as $\lambda \to 0$ and $N\lambda^{2/r} \to \infty$,*

$$\begin{aligned}(V + \lambda J)(\hat{\eta} - \eta_0) &= O_p(N^{-1}\lambda^{-1/r} + \lambda^p),\\ \mathrm{SKL}(\eta_0, \hat{\eta}) &= O_p(N^{-1}\lambda^{-1/r} + \lambda^p), \text{ and}\\ (V + \lambda J)(\hat{\eta}^* - \eta_0) &= O_p(N^{-1}\lambda^{-1/r} + \lambda^p),\\ \mathrm{SKL}(\eta_0, \hat{\eta}^*) &= O_p(N^{-1}\lambda^{-1/r} + \lambda^p),\end{aligned}$$

where constants r and p are defined as above.

The technical proof is put in Appendix B. In the proof, we further assume that the gap times for each subject are distinct, i.e., the gap time counting processes satisfy $N_{ij}(t)N_{ij'}(t) \equiv 0$ for any i and $j \neq j'$.

3 Empirical studies

In this section, we shall carry out some numerical experiments to evaluate the proposed methods. In all the simulations below, $\mathcal{W}(a, b)$ denotes the Weibull distribution with density function $f(t) = \frac{a}{b^a} t^{a-1} e^{-(t/b)^a}$ and thus log hazard function $\eta(t) = \log a + (a-1)\log t - a \log b$.

3.1 Performance of cross validation

To gauge the performance of the cross validation score developed in Sect. 2.3, we tested the method on data with gap times generated from $\mathcal{W}(1.5, 0.7)$ and censoring time from Uniform $(0, 3.5)$. For sample sizes $n = 100, 200$, one hundred replicates each were generated and cubic splines in Example 1 were calculated with λ on a grid $\log_{10}(\lambda) = (-8)(0.1)(-2)$. This grid is wide enough to cover most common choices of λ. For a fixed λ, the Kullback–Leibler loss

$$\mathrm{KL}(\eta_0, \eta_\lambda) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{\Xi_i} \int_0^{X_{ij}} \left\{ (\eta_0(t) - \eta_\lambda(t)) e^{\eta_0(t)} - (e^{\eta_0(t)} - e^{\eta_\lambda(t)}) \right\} dt, \quad (13)$$

represents the closeness between the estimate η_λ and the true η_0. With a varying λ, (13) defines a set of losses $\mathrm{KL}(\eta_0, \eta_\lambda)$. The minimum in the set is thus the minimum attainable loss through the minimization of the penalized likelihood (2). Hence the corresponding smoothing parameter λ is the best one. To evaluate the performance of a cross-validation score $V_\alpha(\lambda)$ in smoothing parameter selection, we first identified the λ that minimizes $V_\alpha(\lambda)$, computed the loss $\mathrm{KL}(\eta_0, \eta_\lambda)$ of the corresponding estimate η_λ, and then compared it with the minimum attainable loss. These comparisons for $\alpha = 1, 1.2, 1.4, 1.6, 1.8$ are summarized in Fig. 1. Plotted in the left and center frames are the $\mathrm{KL}(\eta_0, \eta_\lambda)$ corresponding to the modified ($\alpha = 1.4$) and unmodified ($\alpha = 1$) CV scores for sample sizes $n = 100, 200$ versus the minimum attainable $\mathrm{KL}(\eta_0, \eta_\lambda)$ on the grid. Clearly, the wild failures of $V_\alpha(\lambda)$ were effectively reduced with $\alpha = 1.4$.

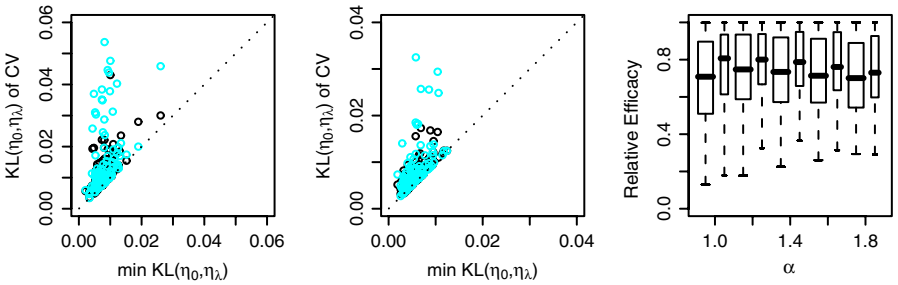

Fig. 1 Performance of cross-validation scores $V_\alpha(\lambda)$. *Left*: Loss achieved by $V_\alpha(\lambda)$ with $\alpha = 1$ (faded) and $\alpha = 1.4$ for $n = 100$. *Center*: Loss achieved by $V_\alpha(\lambda)$ with $\alpha = 1$ (faded) and $\alpha = 1.4$ for $n = 200$. *Right*: Relative efficacy of $V_\alpha(\lambda)$ with $\alpha = 1, 1.2, 1.4, 1.6, 1.8$, for $n = 100$ (fatter boxes) and $n = 200$ (thinner boxes)

The relative efficacies of all the five CV scores in both simulations, defined as the ratios of the minimum attainable $\mathrm{KL}(\eta_0, \eta_\lambda)$ to the $\mathrm{KL}(\eta_0, \eta_\lambda)$ corresponding to the CV scores, are shown in the right frame in box plots. These box plots show that the best performance of $V_\alpha(\lambda)$ was achieved with α in the range $1.2 \sim 1.4$. Our empirical studies (not reported here) with other distribution configurations also suggest an α value in this range.

3.2 Coverage properties

We now assess the coverage properties of the point-wise intervals derived in Sect. 2.4. We simulated one hundred data replicates mimicking the bladder tumor data to be studied in Sect. 4. Each data replicate has $n = 150$ simulated patients. Each tumor recurrence is associated with three covariate $(U_{\langle t\rangle}, U_{\langle n\rangle}, U_{\langle r\rangle})$, representing respectively the treatment, the number of initial tumors, and the number of up-to-date recurrences. $U_{\langle t\rangle}$ was simulated to take value 0 or 1 with equal probability, and $U_{\langle n\rangle}$ took value from 1 to 8 with probability weights (50, 11, 10, 4, 5, 2, 0, 3), which are the actual counts of the patients with the specified numbers of initial tumors in the bladder tumor study. Given $U_{\langle t\rangle} = u_{\langle t\rangle}, U_{\langle n\rangle} = u_{\langle n\rangle}$, and the number $u_{\langle r\rangle}$ of past recurrences, the tumor recurrence time T was generated from $\mathcal{W}(a(u), b(u))$ with

$$a(u) = (0.3 + 0.05u_{\langle r\rangle})^{1/(1+u_{\langle t\rangle})} \quad \text{and} \quad b(u) = 50/\sqrt{1 + u_{\langle n\rangle}}.$$

Note that under this setting, the log hazard function $\eta(t, u)$ is non-additive in $u_{\langle t\rangle}$ and $u_{\langle r\rangle}$ but additive in $u_{\langle n\rangle}$. The censoring time was generated from Uniform(30, 60). For each data replicate, the log hazard was estimated using the tensor product splines in Example 2 and the model (15) to be applied to the bladder tumor study. The smoothing parameters were selected using $\alpha = 1.4$.

To examine the point-wise coverage, we first fixed the covariates at $(u_{\langle t\rangle}, u_{\langle r\rangle}, u_{\langle n\rangle}) = (0, 1, 1)$. Then for each time point t on the grid (5)(0.5)(50), we computed the empirical coverage of the true log hazard η at the point $(t, (0, 1, 1))$ as the percentage of the 100 replicates whose 95% confidence interval estimates cover the true value $\eta(t, (0, 1, 1))$. The empirical point-wise coverage is then plotted in the left frame of Fig. 2. We can see the coverage starts low when t is small (<10), which is expected since there are only a small number of recurrence times falling below 10. As t increase, the point-wise coverage goes up and stays close to the nominal value 0.95. To assess the width of the confidence intervals, we selected two replicates that have the lowest and the highest across-replicate average coverage (= percentage of the grid points where the interval estimates from the data replicate cover the true values). Then the hazard fits (solid lines) for the two replicates are plotted, together with the true hazard (dotted lines) and the connected 95% point-wise confidence intervals (dashed lines), in the middle (worst) and the right (best) frames of Fig. 2. Both plots show a proper scale for the fit and the interval estimates. We also tried the simulations on covariate settings $(u_{\langle t\rangle}, u_{\langle r\rangle}, u_{\langle n\rangle}) = (0, 3, 1), (0, 1, 3), (0, 3, 3), (1, 1, 1), (1, 3, 1), (1, 1, 3), (1, 3, 3)$. The results, though not shown here, were similar.

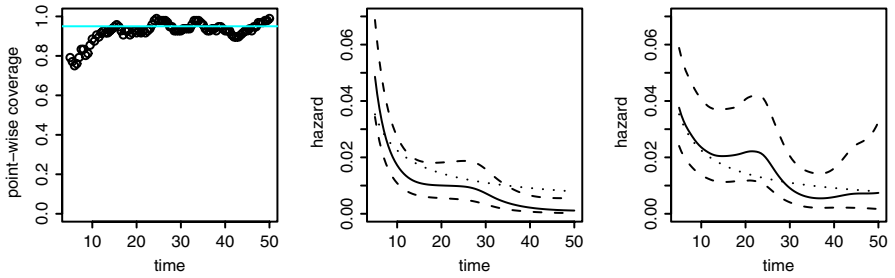

Fig. 2 Coverage property of Bayesian confidence intervals. *Left*: Point-wise coverage over 100 data replicates. The faded line is the nominal level 0.95. *Center and right*: Hazard fits and confidence intervals for the replicates with the lowest (center) and the highest (right) coverage. The solid lines are the fits, the dashed lines 95% confidence intervals, and the dotted lines the test hazard

4 Application: bladder tumor cancer data

In this section, we shall illustrate the proposed methods using the well-known bladder tumor cancer data reported by Byar (1980). The data were from a randomized clinical trial conducted by the Veterans Administration Co-operative Urological Group between 1971 and 1976 and consisted of 118 patients with superficial bladder tumors. The tumors were removed transurethrally, and patients were then randomly assigned to one of three treatments: placebo, pyridoxine or thiotepa. Patients were examined every 3 months for recurrence of tumor and any new tumors were removed. The available data are for the placebo and the thiotepa groups. There were 47 patients in the placebo group with a total of 87 observed recurrences and 38 patients in the thiotepa group with a total of 45 observed recurrences. Besides the treatment and number of months to the event since last tumor occurrence, the number of initial tumors and diameter of the largest initial tumor were also recorded for each patient. Particularly, the number of initial tumors ranged from 1 to 8 with the respective counts of patients equal to (50, 11, 10, 4, 5, 2, 0, 3). In our analysis, we will also include the number of past tumor recurrences as a covariate for each gap time.

Let $\eta(t, u)$ be the log hazard for gap times between tumor recurrences of a patient when t months has passed since the last tumor occurrence, given that the covariate value of the patient is u. Here $u = (u_{\langle t\rangle}, u_{\langle n\rangle}, u_{\langle r\rangle}, u_{\langle s\rangle})$ consists of the treatment $u_{\langle t\rangle}$ as a categorical variable with two levels (placebo or thiotepa), the number of initial tumors $u_{\langle n\rangle}$, the number of past tumor recurrences $u_{\langle r\rangle}$ and the size of the largest initial tumor in diameter $u_{\langle s\rangle}$ as continuous variables. The tensor product splines in Example 2 were used in our following analysis. We first fitted a model with all the two-way gap-time-covariate interactions,

$$\begin{aligned}\eta(t_{ij}, u_{ij}) = {} & \eta_\emptyset + \eta_1(t_{ij}) + \eta_2(u_{i\langle t\rangle}) + \eta_3(u_{i\langle n\rangle}) + \eta_4(u_{ij\langle r\rangle}) + \eta_5(u_{i\langle s\rangle}) \\ & + \eta_{1,2}(t_{ij}, u_{i\langle t\rangle}) + \eta_{1,3}(t_{ij}, u_{i\langle n\rangle}) + \eta_{1,4}(t_{ij}, u_{ij\langle r\rangle}) \\ & + \eta_{1,5}(t_{ij}, u_{i\langle s\rangle}), \end{aligned} \tag{14}$$

where i is the index for subject and j is the index for event. The model was fitted with $\alpha = 1.4$ for smoothing parameter selection. After applying the model selection tool

developed in Gu (2004) for penalized likelihood estimation, the terms $\eta_5(u_{i\langle s\rangle})$, $\eta_{1,3}(t_{ij}, u_{i\langle n\rangle})$ and $\eta_{1,5}(t_{ij}, u_{i\langle s\rangle})$ were found to be negligible. Eliminating these terms from (14), our final model is

$$\eta(t_{ij}, u_{ij}) = \eta_\emptyset + \eta_1(t_{ij}) + \eta_2(u_{i\langle t\rangle}) + \eta_3(u_{i\langle n\rangle}) + \eta_4(u_{ij\langle r\rangle}) + \eta_{1,2}(t_{ij}, u_{i\langle t\rangle}) + \eta_{1,4}(t_{ij}, u_{ij\langle r\rangle}). \tag{15}$$

First, note that both of the terms involving $u_{\langle s\rangle}$ in (14) are not included in the final model (15). It indicates that the size of the largest initial tumor is not associated with tumor recurrences. This is a common finding in many of the previous studies on the data. Second, model (15) indicates a violation of hazard proportionality. The significance of the two interactions $\eta_{1,2}(t_{ij}, u_{i\langle t\rangle})$ and $\eta_{1,4}(t_{ij}, u_{ij\langle r\rangle})$ suggests that the effects of treatment and number of recurrences experienced vary with gap time.

Then with model (15) fitted, Bayesian confidence intervals can be constructed by the formula in Sect. 2.4. Plotted in Fig. 3 are 8 slices of the hazard fit along with their 95% Bayesian confidence intervals, at the covariate level combinations $(u_{\langle t\rangle}, u_{\langle n\rangle}, u_{\langle r\rangle}) =$ ("p", 1, 1), ("p", 1, 3), ("p", 3, 1), ("p", 3, 3), ("t", 1, 1), ("t", 1, 3), ("t", 3, 1), ("t", 3, 3), where $u_{\langle t\rangle}$ = "p" and $u_{\langle t\rangle}$ = "t" correspond to placebo and thiotepa respectively, and 1 and 3 are the lower and upper quartiles of $u_{\langle n\rangle}$ and $u_{\langle r\rangle}$ in the data. An observation from these plot is that when fixing other covariate components at the specified levels, a patient treated with placebo has higher tumor recurrence hazard than one treated with thiotepa, a patient with three initial tumors has higher tumor recurrence hazard than one with only one initial tumor, and a patient with three past tumor recurrences has higher tumor recurrence hazard than one with only one tumor recurrence.

For the record, the final model (15) took 7.7 CPU minutes to compute, on a Dell PowerEdge 2950 workstation with dual Xeon dual core 3.0 GHz CPUs and 16 GB RAMs running openSUSE 11.0 and R 2.7.1.

5 Discussion

In this article, we have proposed a nonparametric penalized likelihood approach to estimating the gap time hazard function in recurrent event data. Delete-one-subject cross-validation is used to select smoothing parameter. Point-wise confidence intervals of hazard function are developed for inference purpose. Asymptotic convergence rates of the estimates are also established. The empirical studies have evaluated the effectiveness of smoothing parameter selection and the coverage property of the proposed confidence intervals. An interactive analysis of the well-known bladder tumor cancer data demonstrates the usefulness of the proposed techniques.

In many studies, the covariate is associated only with each subject. This article only focuses on processes on gap time domain. Thus the proposed techniques allow a covariate U that changes at (and only at) each recurrence of a subject. This can clearly apply to more general cases when diagnostic measurements are taken at each recurrence. Another possible application of this generalization can be letting U_{ij} represent part of the history for subject i up to recurrence j. For example, let U_{ij} be the number

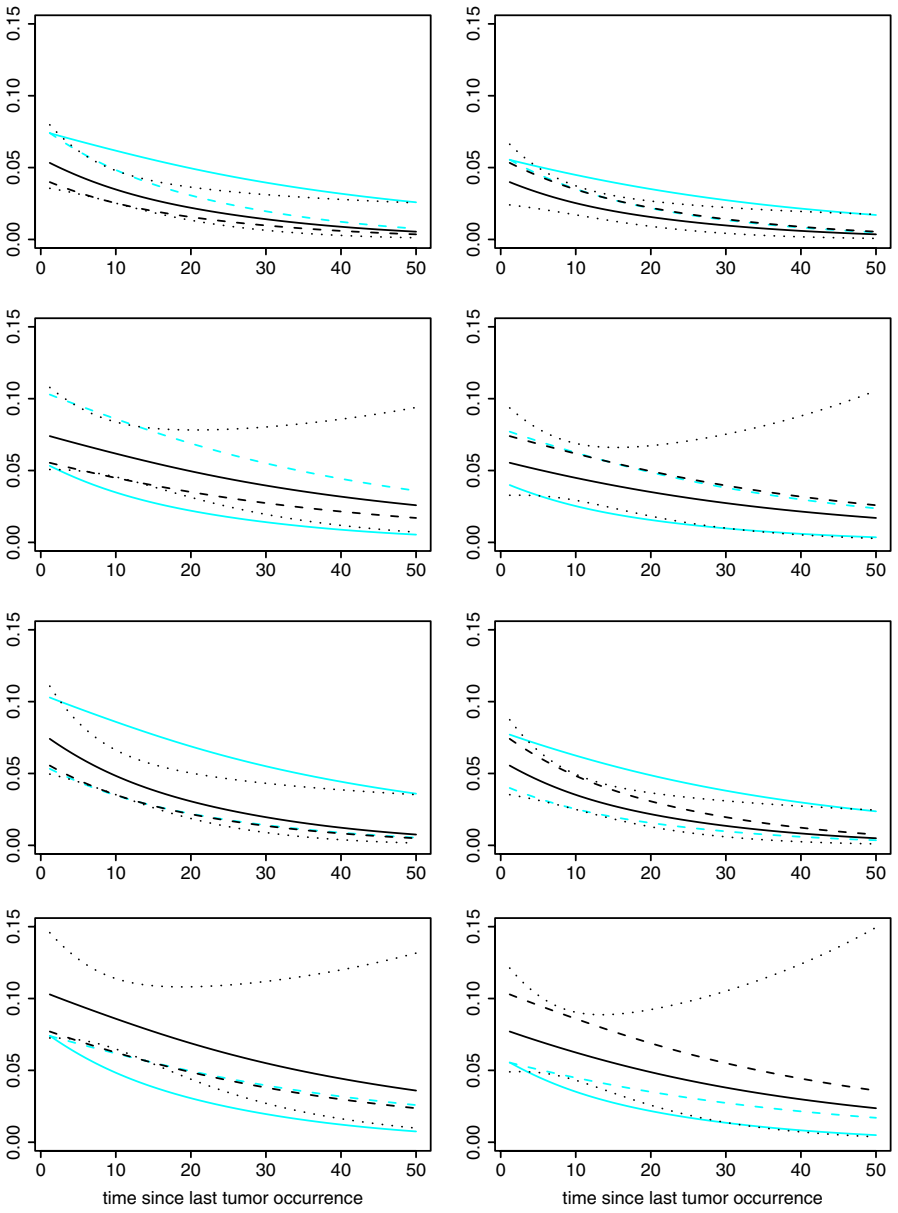

Fig. 3 Hazard estimates for bladder tumor cancer data. *Left*: Placebo ($u_{\langle t \rangle}$ = "p"). *Right*: Thiotepa ($u_{\langle t \rangle}$ = "t"). From top to bottom: $(u_{\langle n \rangle}, u_{\langle r \rangle}) = (1, 1), (1, 3), (3, 1), (3, 3)$. The hazard fit is in solid line and the 95% Bayesian confidence intervals are in dotted lines; superimposed are the other hazard fits whose covariate values differ from those of the solid line only in $u_{\langle t \rangle}$ (dashed line), $u_{\langle n \rangle}$ (faded dashed line), or $u_{\langle r \rangle}$ (faded solid line)

of recurrences experienced by subject i up to recurrence j as in our analysis of the bladder tumor cancer data.

For simplicity, in this article we only consider independent right censoring at the last event. However, in some situations each recurrence can subject to left truncation and right censoring. An example of such recurrent event data appeared in an AIDS study on the effect of drug ribavirin described in Wei et al. (1989), where each of the three potential events for a patient could be right censored. A more common situation is that the first recurrence can be left truncated. In these situations, one can simply modify the proposed method by replacing $(0, X_{ij})$ with the corresponding left truncation and right censoring times in (2).

Acknowledgements The author is grateful to his advisor Chong Gu and the referees for their helpful comments that have significantly improved the paper.

Appendix A: derivation of cross-validation score

In this section, we shall derive the cross-validation score (9) from the Kullback–Leibler distance (8). Dropping the terms that do not involve η_λ, one can estimate the remaining part of (8) by

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_0^{X_{ij}} e^{\eta_\lambda(t,U_{ij})}dt - \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_0^{X_{ij}} \eta_\lambda(t,U_{ij})e^{\eta_0(t,U_{ij})}dt. \tag{16}$$

The first term of (16) is directly computable. But the second term $\mu_{\eta_0}(\eta_\lambda)$ involves the unknown $\eta_0(t,u)$ and has to be estimated.

Let $N(t) = I_{[X\le t,\delta=1]}$ be the event process and $Y(t) = I_{[X\ge t]}$ the at-risk process. Write $A(t) = \int_0^t Y(s)e^{\eta_0(s,U)}ds$ and $M(t) = N(t) - A(t)$. Although $M(t)$ is not a martingale due to the lack of a filtration, Gill (1980), (27) showed that like a martingale,

$$E\left[\int_{\mathcal{T}} f(s)dM(s)\right] = 0 \tag{17}$$

for a predictable function $f(t)$. Then for $f(t,u)$ continuous in t, $\forall u \in \mathcal{U}$, and independent of X, $E[\int_{\mathcal{T}} f(t,U)dM(t)] = 0$. "Estimating" 0 by the sample mean $N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{K_i+1}\int_{\mathcal{T}} f(t,U_{ij})dM_{ij}(t)$, one has

$$\begin{aligned} 0 &\approx \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{K_i+1}\left\{\int_{\mathcal{T}} f(t,U_{ij})dN_{ij}(t) - \int_{\mathcal{T}} f(t,U_{ij})Y_{ij}(t)e^{\eta_0(t,U_{ij})}dt\right\} \\ &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{K_i+1}\left\{\delta_{ij}f(X_{ij},U_{ij}) - \int_0^{X_{ij}} f(t,U_{ij})e^{\eta_0(t,U_{ij})}dt\right\}, \end{aligned} \tag{18}$$

Setting $f(t, U_{ij}) = \eta^{[i]}_{\lambda,\tilde{\eta}}(t, U_{ij})$ yields an estimate

$$\tilde{\mu}_{\eta_0}(\eta_\lambda) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{K_i+1}\int_0^{X_{ij}} \eta^{[i]}_{\lambda,\tilde{\eta}}(t, U_{ij})e^{\eta_0(t,U_{ij})}dt \approx \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\eta^{[i]}_{\lambda,\tilde{\eta}}(X_{ij}, U_{ij}), \tag{19}$$

of $\mu_{\eta_0}(\eta_\lambda)$ in (18), where $\eta^{[i]}_{\lambda,\tilde{\eta}}$ minimizes the delete-one-subject version of the quadratic approximation of (2) at $\tilde{\eta} = \eta_\lambda$:

$$-\frac{1}{N-K_i-1}\sum_{\substack{l=1\\ l\neq i}}^{n}\sum_{j=1}^{K_l}\eta(X_{lj}, U_{lj}) + \mu_{\tilde{\eta}}(\eta) - V_{\tilde{\eta}}(\tilde{\eta}, \eta) + \frac{1}{2}V_{\tilde{\eta}}(\eta, \eta) + \frac{\lambda}{2}J(\eta). \tag{20}$$

Set $\tilde{\eta} = \eta_\lambda$ and write $\boldsymbol{\psi} = (\boldsymbol{\xi}^T, \boldsymbol{\phi}^T)^T$ and $\check{\mathbf{c}} = (\mathbf{c}^T, \mathbf{d}^T)^T$. Rewrite (7) as $H\check{\mathbf{c}} = \check{Q}\mathbf{1}/N + \mathbf{g}$, where $H = V_{\tilde{\eta}}(\boldsymbol{\psi}, \boldsymbol{\psi}^T) + \mathrm{diag}(\lambda Q, 0)$, $\check{Q} = (\sum_{j=1}^{K_1}\boldsymbol{\psi}(X_{1j}, U_{1j}), \ldots, \sum_{j=1}^{K_n}\boldsymbol{\psi}(X_{nj}, U_{nj}))$, and $\mathbf{g} = V_{\tilde{\eta}}(\boldsymbol{\psi}, \tilde{\eta}) - \mu_{\tilde{\eta}}(\boldsymbol{\psi})$. Similarly, the coefficient vector of $\eta^{[i]}_{\lambda,\tilde{\eta}}$, the minimizer of (20), is

$$\begin{aligned}\check{\mathbf{c}}^{[i]} &= H^{-1}\left(\frac{\check{Q}\mathbf{1} - \sum_{j=1}^{K_i}\boldsymbol{\psi}(X_{ij}, U_{ij})}{N-K_i-1} + \mathbf{g}\right)\\ &= \check{\mathbf{c}} - \frac{1}{N}H^{-1}\check{Q}\mathbf{1} + \frac{1}{N-K_i-1}H^{-1}\check{Q}\mathbf{1}\\ &\quad - \frac{1}{N-K_i-1}H^{-1}\left(\sum_{j=1}^{K_i}\boldsymbol{\psi}(X_{ij}, U_{ij})\right).\end{aligned} \tag{21}$$

Note that $\eta^{[i]}_{\lambda,\tilde{\eta}}(X_{ij}, U_{ij}) = \boldsymbol{\psi}(X_{ij}, U_{ij})^T\check{\mathbf{c}}^{[i]}$. Combining (16), (19) and (21) yields the unadjusted cross-validation score

$$\begin{aligned}&-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\left\{\delta_{ij}\eta_\lambda(X_{ij}, U_{ij}) - \int_0^{X_{ij}} e^{\eta_\lambda(t,U_{ij})}dt\right\}\\ &+\frac{1}{N}\left\{\mathbf{1}^T(I/N - D)\check{Q}^T H^{-1}\check{Q}\mathbf{1} + \mathrm{tr}(D\check{Q}^T H^{-1}\check{Q})\right\}.\end{aligned} \tag{22}$$

Appendix B: proof of asymptotic convergence

In this section, we shall prove the asymptotic convergence rates of the estimates in three steps through an eigenvalue analysis. Let $\hat{\eta}$ and $\hat{\eta}^*$ be the minimizers of (2) in $\mathcal{H}$ and $\mathcal{H}_n$.

Write the gap time processes as $N_{ij}(t) = I_{[X_{ij}\le t,\delta_{ij}=1]}$, $Y_{ij}(t) = I_{[X_{ij}\ge t]}$, and $A_{ij}(t) = \int_0^t Y_{ij}(s)e^{\eta_0(s,U_{ij})}ds$. Although $M_{ij}(t) = N_{ij}(t) - A_{ij}(t)$ is not a martingale, it has a martingale-type property as in (17). The penalized likelihood functional (2) can be written as

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\left\{\int_{\mathcal{T}}\eta_{ij}dN_{ij}(t) - \int_{\mathcal{T}}Y_{ij}e^{\eta_{ij}}dt\right\} + \frac{\lambda}{2}J(\eta), \tag{23}$$

where $\eta_{ij}(t) = \eta(t, U_{ij})$.

From Condition 1, there exist eigenvalues π_ν (as in Condition 2) and the associated eigenfunctions ζ_ν such that $V(\zeta_\nu, \zeta_\mu) = \delta_{\nu,\mu}$ and $J(\zeta_\nu, \zeta_\mu) = \pi_\nu\delta_{\nu,\mu}$, where $\delta_{\nu,\mu}$ is the Kronecker delta. And any function f satisfying $J(f) < \infty$ can be expressed as a Fourier series expansion $f = \sum_\nu f_\nu\zeta_\nu$ with the Fourier coefficients $f_\nu = V(f, \zeta_\nu)$. See, e.g., Weinberger (1974) and Gu (2002). Simple calculus manipulations then yield the following lemma.

Lemma 1 *Under Condition 2, as* $\lambda \to 0$, *the sums* $\sum_\nu \frac{\lambda\pi_\nu}{(1+\lambda\pi_\nu)^2}$, $\sum_\nu \frac{1}{(1+\lambda\pi_\nu)^2}$, *and* $\sum_\nu \frac{1}{1+\lambda\pi_\nu}$ *are all of order* $O(\lambda^{-1/r})$.

Step 1: Linear Approximation.

A linear approximation $\tilde\eta$ to $\hat\eta$ is the minimizer of a quadratic approximation to (2),

$$-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\left\{\int_{\mathcal{T}}\eta_{ij}dN_{ij}(t) - \int_{\mathcal{T}}\eta_{ij}Y_{ij}e^{\eta_{0,ij}}dt\right\} + \frac{1}{2}V(\eta-\eta_0) + \frac{\lambda}{2}J(\eta), \tag{24}$$

after dropping the terms not involving η. Plugging the Fourier series expansions $\eta = \sum_\nu \eta_\nu\zeta_\nu$ and $\eta_0 = \sum_\nu \eta_{\nu,0}\zeta_\nu$ in (24), one obtains the Fourier coefficients

$$\tilde\eta_\nu = (\beta_\nu + \eta_{\nu,0})/(1+\lambda\pi_\nu)$$

of $\tilde\eta$, where $\beta_\nu = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}}\zeta_{\nu,ij}dM_{ij}(t)$ with $\zeta_{\nu,ij}(t) = \zeta_\nu(t, U_{ij})$. Further assuming that gap times for each subject are distinct, one has $N_{ij}(t)N_{ij'}(t) \equiv 0$ for any i and $j \ne j'$. Hence $E[\beta_\nu] = 0$ and $E[\beta_\nu^2] = N^{-1}$, which then give

$$\begin{aligned} E[V(\tilde\eta-\eta_0)] &= \frac{1}{n}\sum_\nu\frac{1}{(1+\lambda\rho_\nu)^2} + \lambda\sum_\nu\frac{\lambda\rho_\nu}{(1+\lambda\rho_\nu)^2}\rho_\nu\eta_{\nu,0}^2,\\ E[\lambda J(\tilde\eta-\eta_0)] &= \frac{1}{n}\sum_\nu\frac{\lambda\rho_\nu}{(1+\lambda\rho_\nu)^2} + \lambda\sum_\nu\frac{(\lambda\rho_\nu)^2}{(1+\lambda\rho_\nu)^2}\rho_\nu\eta_{\nu,0}^2. \end{aligned} \tag{25}$$

Combining Lemma 1 and (25), we obtain that $(V+\lambda J)(\tilde\eta-\eta_0) = O_p(N^{-1}\lambda^{-1/r} + \lambda^p)$, as $N \to \infty$ and $\lambda \to 0$.

Step 2: Approximation Error.

We now investigate the approximation error $\hat\eta - \tilde\eta$ and prove the convergence rate of $\hat\eta$. Define $A_{f,g}(\alpha)$ and $B_{f,g}(\alpha)$ respectively as the resulting functionals from setting

$\eta = f + \alpha g$ in (23) and (24). Differentiating them with respect to α and then setting $\alpha = 0$ yields

$$\dot{A}_{f,g}(0) = -\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\left\{\int_{\mathcal{T}} g_{ij}dN_{ij}(t) - \int_{\mathcal{T}} g_{ij}Y_{ij}e^{f_{ij}}dt\right\} + \lambda J(f,g), \quad (26)$$

$$\dot{B}_{f,g}(0) = -\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\left\{\int_{\mathcal{T}} g_{ij}dN_{ij}(t) - \int_{\mathcal{T}} g_{ij}Y_{ij}e^{\eta_{0,ij}}dt\right\} + V(f-\eta_0, g) + \lambda J(f,g). \quad (27)$$

Set $f = \hat{\eta}$ and $g = \hat{\eta} - \tilde{\eta}$ in (26), and set $f = \tilde{\eta}$ and $g = \hat{\eta} - \tilde{\eta}$ in (27). Then subtracting the resulted equations gives

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}}(\hat{\eta}-\tilde{\eta})_{ij}(e^{\hat{\eta}} - e^{\tilde{\eta}})_{ij}Y_{ij}dt + \lambda J(\hat{\eta}-\tilde{\eta})$$
$$= V(\tilde{\eta}-\eta_0, \hat{\eta}-\tilde{\eta}) - \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}}(\hat{\eta}-\tilde{\eta})_{ij}(e^{\tilde{\eta}} - e^{\eta_0})_{ij}Y_{ij}dt. \quad (28)$$

The following lemma is needed to proceed.

Lemma 2 *Under Conditions 1, 2, and 4, as* $\lambda \to 0$ *and* $N\lambda^{2/r} \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}} f_{ij}g_{ij}e^{\eta_{0,ij}}Y_{ij}dt = V(f,g) + o_p(\{(V+\lambda J)(f)(V+\lambda J)(g)\}^{1/2}),$$

where $f_{ij} = f(t, U_{ij})$ *and* $g_{ij} = g(t, U_{ij})$.

Proof First, repeated uses of the Cauchy–Schwartz inequality can easily show that $E\left[\left\{\int_{\mathcal{T}}\zeta_\nu\zeta_\mu e^{\eta_0}Ydt - V(\zeta_\nu,\zeta_\mu)\right\}^2\right] < \infty$. Let $f = \sum_\nu f_\nu\zeta_\nu$ and $g = \sum_\mu g_\mu\zeta_\mu$ be the Fourier series expansion of f and g. Then from the Cauchy–Schwartz inequality and Lemma 1,

$$\left|\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}} f_{ij}g_{ij}e^{\eta_{0,ij}}Y_{ij}dt - V(f,g)\right|$$
$$= \left|\sum_\nu\sum_\mu f_\nu g_\mu\left\{\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}}\zeta_{\nu,ij}\zeta_{\mu,ij}e^{\eta_{0,ij}}Y_{ij}dt - V(\zeta_\nu,\zeta_\mu)\right\}\right|$$
$$\le \left\{\sum_\nu\sum_\mu \frac{1}{1+\lambda\rho_\nu}\frac{1}{1+\lambda\rho_\mu}\right.$$

$$\times \left\{ \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{\Xi_i} \int_{\mathcal{T}} \zeta_{v,ij} \zeta_{\mu,ij} e^{\eta_{0,ij}} Y_{ij} dt - V(\zeta_v, \zeta_\mu) \right\}^2 \Bigg\}^{1/2}$$

$$\times \left\{ \sum_{v} \sum_{\mu} (1 + \lambda\rho_v)(1 + \lambda\rho_\mu) f_v^2 g_\mu^2 \right\}^{1/2}$$

$$= O_p(N^{-1/2}\lambda^{-1/r})\{(V + \lambda J)(f)(V + \lambda J)(g)\}^{1/2}. \qquad \square$$

By the mean value theorem, Condition 3, Lemma 2, and (28),

$$\begin{aligned}(c_1 V + \lambda J)(\hat{\eta} - \tilde{\eta})(1 + o_p(1)) \\ \le \{(|1 - c|V + \lambda J)(\hat{\eta} - \tilde{\eta})\}^{1/2} O_p(\{(|1 - c|V + \lambda J)(\tilde{\eta} - \eta_0)\}^{1/2})\end{aligned}$$

for some $c \in [c_1, c_2]$. Then the convergence rate of $\hat{\eta}$ in Theorem 1 follows from that of $\tilde{\eta}$ proved in the previous step.

Step 3: Semiparametric Approximation.

Our last goal is the convergence rate for the minimizer $\hat{\eta}^*$ in the space $\mathcal{H}_n = \mathcal{N}_J \oplus \text{span}\{R_J((X_{ij}, U_{ij}), \cdot) : j = 1, \ldots, K_i; i = 1, \ldots, n\}$.

For any $h \in \mathcal{H} \ominus \mathcal{H}_n$, one has $\delta_{ij} h(X_{ij}, U_{ij}) = \delta_{ij} J(R_J((X_{ij}, U_{ij}), \cdot), h) = 0$, so $\sum_{i=1}^{n} \sum_{j=1}^{\Xi_i} \int_{\mathcal{T}} h_{ij}^2 dN_{ij}(t) = \sum_{i=1}^{n} \sum_{j=1}^{\Xi_i} \delta_{ij} h^2(X_{ij}, U_{ij}) = 0$, where $h_{ij}(t) = h(t, U_{ij})$. Hence, by the same arguments used in the proof of Lemma 2,

$$\begin{aligned} V(h) &= \left| \frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{T}} h_i^2 dN_i(t) - V(h) \right| \\ &= O_p(n^{-1/2}\lambda^{-1/r})(V + \lambda J)(h) = o_p(\lambda J(h)). \end{aligned} \tag{29}$$

Let η^* be the projection of $\hat{\eta}$ in $\mathcal{H}_n$. Setting $f = \hat{\eta}$ and $g = \hat{\eta} - \eta^*$ in (26), some algebra yields

$$\begin{aligned} \lambda J(\hat{\eta} - \eta^*) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{\Xi_i} \int_{\mathcal{T}} (\hat{\eta} - \eta^*)_{ij} dM_{ij}(t) \\ - \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{\Xi_i} \int_{\mathcal{T}} (\hat{\eta} - \eta^*)_{ij} (e^{\hat{\eta}} - e^{\eta_0})_{ij} Y_{ij} dt; \end{aligned} \tag{30}$$

Note that $J(\eta^*, \hat{\eta} - \eta^*) = 0$ and $\beta_v = N^{-1} \sum_{i=1}^{n} \sum_{j=1}^{\Xi_i} \int_{\mathcal{T}} \phi_{v,ij} dM_{ij}(t)$. An application of the Cauchy–Schwartz inequality shows that the first sum in (30) is of order $\{(V + \lambda J)(\hat{\eta} - \eta^*)\}^{1/2} O_p(N^{-1/2}\lambda^{-1/2r})$. By the mean value theorem, Condition 3, Lemma 2, and (29), the second sum in (30) is of order $o_p(\{\lambda J(\hat{\eta} - \eta^*)(V + \lambda J)(\hat{\eta} - \eta_0)\}^{1/2})$. These, combined with (30) and the convergence rates of $\hat{\eta}$, yield $\lambda J(\hat{\eta} - \eta^*) = O_p(N^{-1}\lambda^{-1/r} + \lambda^p)$ and $V(\hat{\eta} - \eta^*) = o_p(N^{-1}\lambda^{-1/r} + \lambda^p)$.

Set $f = \hat{\eta}^*$ and $g = \hat{\eta}^* - \eta^*$ in (26), and set $f = \hat{\eta}$ and $g = \hat{\eta} - \hat{\eta}^*$ in (26). Subtracting the resulted equations yields

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}}(\hat{\eta}^* - \eta^*)_{ij}(e^{\hat{\eta}^*} - e^{\eta^*})_{ij}Y_{ij}dt + \lambda J(\hat{\eta}^* - \eta^*) + \lambda J(\hat{\eta} - \eta^*) \\
&\quad = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}}(\hat{\eta} - \eta^*)_{ij}dM_{ij}(t) + \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}}(\hat{\eta} - \eta^*)_{ij}(e^{\eta_0} - e^{\hat{\eta}})_{ij}Y_{ij}dt \\
&\qquad + \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{\Xi_i}\int_{\mathcal{T}}(\hat{\eta}^* - \eta^*)_{ij}(e^{\hat{\eta}} - e^{\eta^*})_{ij}Y_{ij}dt. \qquad (31)
\end{aligned}$$

By the mean value theorem, Condition 3, and Lemma 2, the left-hand side of (31) is bounded from below by

$$(c_1 V + \lambda J)(\hat{\eta}^* - \eta^*)(1 + o_p(1)) + \lambda J(\hat{\eta} - \eta^*).$$

For the right-hand side, the first and second terms are of the order $O_p(N^{-1}\lambda^{-1/r} + \lambda^p)$ by similar arguments for (30), and the third term is of the order

$$\{(V + \lambda J)(\hat{\eta}^* - \eta^*)\}^{1/2} o_p(\{\lambda J(\hat{\eta} - \eta^*)\}^{1/2})$$

by Condition 3, Lemma 2, and (29). Putting things together, one obtains $(V + \lambda J)(\hat{\eta}^* - \eta^*) = O_p(N^{-1}\lambda^{-1/r} + \lambda^p)$. Then the convergence rate of $\hat{\eta}^*$ in Theorem 1 follows.

References

Aalen OO (1988) Heterogeneity in survival analysis. Stat Med 7:1121–1137

Aalen O, Husebye E (1991) Statistical analysis of repeated events forming renewal processes. Stat Med 10: 1227–1240

Aronszajn N (1950) Theory of reproducing kernels. Trans Am Math Soc 68:337–404

Byar DP (1980) The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo pyridoxine, and topical thiotepa. In: Pavone-Macaluso M, Smith PH, Edsmyn F (eds) Bladder tumors and other topics in urological oncology. Plenum, New York, pp 363–370

Du P, Gu C (2006) Penalized likelihood hazard estimation: efficient approximation and Bayesian confidence intervals. Stat Probab Lett 76:244–254

Gail MH, Santner TJ, Brown CC (1980) An analysis of comparative carcinogenisis experiments based on multiple times to tumor. Biometrics 36:255–266

Gill RD (1980) Nonparametric estimation based on censored observations of a Marcov renewal process. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 53:97–116

Gu C (1996) Penalized likelihood hazard estimation: a general procedure. Stat Sin 6:861–876

Gu C (2002) Smoothing spline ANOVA models. Springer, New York

Gu C (2004) Model diagnostics for smoothing spline ANOVA models. Can J Stat 32:347–358

Gu C, Ma P (2005a) Generalized nonparametric mixed-effect models. J Comput Graph Stat 14:485–504

Gu C, Ma P (2005b) Optimal smoothing in nonparametric mixed-effect models. Ann Stat 33:1357–1379

Hougaard P (2000) Analysis of multivariate survival data. Springer, New York

Joly P, Commenges D, Letenneur L (1998) A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. Biometrics 54:185–194

Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data. Wiley, New York
Karcher P, Wang Y (2001) Generalized nonparametric mixed effects models. J Comput Graph Stat 10:641–655
Lin DY, Sun W, Ying Z (1999) Nonparametric estimation of the gap time distribution for serial events with censored data. Biometrika 86:59–70
Nychka D (1988) Bayesian confidence intervals for smoothing splines. J Am Stat Assoc 83:1134–1143
O'Sullivan F (1988a) Fast computation of fully automated log-density and log-hazard estimators. SIAM J Sci Stat Comput 9:363–379
O'Sullivan F (1988b) Nonparametric estimation of relative risk using splines and cross-validation. SIAM J Sci Stat Comput 9:531–542
Peña E, Strawderman R, Hollander M (2001) Nonparametric estimation with recurrent event data. J Am Stat Assoc 96:1299–1315
Rondeau V, Mathoulin-Pellissier S, Jacqmin-gadda H, Brouste V, Soubeyran P (2007) Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. Biostatistics 8(4):708–721
Strawderman RL (2005) The accelerated gap times model. Biometrika 92:647–666
Vaupel JW, Manton KG, Stallard E (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16:439–454
Wahba G (1983) Bayesian "confidence intervals" for the cross-validated smoothing spline. J R Stat Soc [Ser B] 45:133–150
Wahba G (1990) Spline models for observational data, volume 59 of CBMS-NSF regional conference series in applied mathematics. SIAM, Philadelphia
Wang Y (1998) Mixed-effects smoothing spline ANOVA. J R Stat Soc [Ser B] 60:159–174
Wang M-C, Chang S-H (1999) Nonparametric estimation of a recurrent survival function. J Am Stat Assoc 94:146–153
Wei LJ, Lin DY, Weissfeld L (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc 84:1065–1073
Weinberger HF (1974) Variational methods for eigenvalue approximation, volume 15 of CBMS-NSF regional conference series in applied mathematics. SIAM, Philadelphia
Zeng D, Lin DY (2007) Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). J Roy Stat Soc [Ser B] 69:507–564
Zucker DM, Karr AF (1990) Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. Ann Stat 18:329–353