ASYMPTOTICS FOR HAZARD REGRESSION

by

Young N. Truong and Charles J. Stone

Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2165

August 1996

Asymptotics for Hazard Regression

Young K. Truong*
The University of North Carolina at Chapel Hill

Charles J. Stone†
University of California at Berkeley

August 6, 1996

Abstract

In hazard regression (HARE), the logarithm of the conditional hazard function of a survival time given a covariate is modeled by a sum of polynomial splines and their tensor products. Under appropriate conditions, it has been shown that the (nonadaptive) HARE estimate of the conditional log-hazard function possesses an optimal L_2 rate of convergence. The current paper considers the L_∞ rates of convergence and the distributional properties of HARE estimates of the conditional hazard, cumulative hazard, survival and density functions. In particular, it will be shown that these estimates are asymptotically normal.

*This research was supported in part by National Science Foundation Grant DMS-9403800 and National Institutes of Health Grant CA61937.

†This research was supported in part by National Science Foundation Grant DMS-9504463.

AMS 1991 subject classifications. Primary 62G07; secondary 62G20.

Key words and phrases. Asymptotic normal distribution; Density function; Hazard function; Maximum likelihood; Tensor product of polynomial splines; Survival function.

1 Introduction

Let T and C be nonnegative random variables having a joint distribution that depends on an M-dimensional vector $\mathbf{x} = (x_1, \ldots, x_M)$ of covariates ($M = 0$ when there are no covariates). In survival analysis, T and C are referred to as the survival time (or failure time) and censoring time, respectively. Set $Y = \min(T, C)$ and $\delta = \operatorname{ind}(T \leq C)$. The indicator random variable δ equals 1 if failure occurs on or before the censoring time ($T \leq C$) and it equals 0 otherwise. The observable time Y is said to be uncensored or censored according as $\delta = 1$ or $\delta = 0$. For identifiability, it is assumed that T and C are independent.

Let $f(t|\mathbf{x})$ and $F(t|\mathbf{x})$ denote the density function and distribution function, respectively, of T. The survival, hazard, cumulative hazard and log-hazard functions are defined by

$$S(t|\mathbf{x}) = 1 - F(t|\mathbf{x}), \quad \lambda(t|\mathbf{x}) = f(t|\mathbf{x})/S(t|\mathbf{x}), \quad H(t|\mathbf{x}) = \int_0^t \lambda(u|\mathbf{x})\, du$$

and

$$\phi(t|\mathbf{x}) = \log \lambda(t|\mathbf{x}), \qquad t \geq 0.$$

Let $F_C(z|\mathbf{x})$ denote the distribution function of C, which depends on $\mathbf{x}$, and set $S_C(t|\mathbf{x}) = 1 - F_C(t|\mathbf{x})$.

In the HARE methodology for survival data analysis [see Kooperberg et al. (1995a)], the logarithm of the hazard function of a survival time is approximated by a function ϕ^* having the form of a specified sum of functions of at most d of the variables $t, x_1, \ldots, x_M$. Subject to this form, the approximation is chosen to maximize the expected log-likelihood. Maximum likelihood and sums of tensor products of polynomial splines are then used to construct an estimate $\widehat{\phi}$ of this approximation based on a random sample. The corresponding maximum likelihood estimates $\widehat{\lambda}$, $\widehat{H}$, $\widehat{S}$ and $\widehat{f}$ of approximations of the hazard, cumulative hazard, survival and density functions can be easily derived using the relationships, $\lambda(t|\mathbf{x}) = \exp\, \phi(t|\mathbf{x})$, $H(t|\mathbf{x}) = \int_0^t \lambda(u|\mathbf{x})\, du$, $S(t|\mathbf{x}) = \exp(-H(t|\mathbf{x}))$ and $f(t|\mathbf{x}) = \lambda(t|\mathbf{x})\, S(t|\mathbf{x})$.

Under appropriate conditions, it is shown in Kooperberg et al. (1995b) that $\widehat{\phi}$ possesses an optimal L_2 rate of convergence that depends only on d and a suitably defined smoothness parameter.

A useful feature in the HARE methodology for survival data analysis is that the space G of approximations is chosen adaptively [see Kooperberg et al. (1995a, 1996) and Intrator and Kooperberg (1995)]. Based on our experience in analyzing survival data, we have found the approach to be quite promising. The current paper lends further theoretical support to these methodologies by establishing L_∞ rates of convergence and aymptotic normality for nonadaptive estimates of the hazard, density, survival and cumulative hazard functions.

A log-hazard model is said to be *saturated* if $d = M+1$ and *unsaturated* if $d < M+1$. In establishing the L_∞ rates of convergence and the asymptotic normality of $\widehat{\phi}$, $\widehat{\lambda}$, $\widehat{H}$, $\widehat{S}$ and $\widehat{f}$ for log-hazard models in this paper, we restrict attention to the saturated models. Conceivably, these results also hold for unsaturated models, but extending the L_∞ results in Stone (1989, 1991) to additive and other unsaturated models remains an open problem.

The rest of the paper is organized as follows. Section 2 describes the main results of the paper. Specifically, errors in approximating the log-hazard, hazard, cumulative hazard, survival and density functions are given in Theorem 1. Errors in estimating these approximations are described in Theorem 2. These results indicate that the (nonadaptive) HARE estimates achieve the usual optimal L_2 and L_∞ rates of convergence. Asymptotic normality of these estimates is described in Theorem 3. The proofs of Theorems 1–3 are given in Sections 3–5, respectively. Additional details for the proof of Theorem 1 are given in Section 6.

2 Statements of results

This section describes the main results of the paper. The description involves log-likelihood and expected log-likelihood for censored survival data, polynomial splines and their tensor products, errors of approximation based on the linear space G of tensor products of polynomial splines, maximum likelihood estimates, optimal rates of convergence, asymptotic normality, asymptotic variances, and standard errors. We start with the log-likelihood and expected log-likelihood functions.

2.1 Expected log-likelihood function

The log-likelihood based on $(Y, \delta, \mathbf{x})$ is given by

$$\begin{aligned}\log\{[f(Y|\mathbf{x})]^{\delta}[S(Y|\mathbf{x})]^{1-\delta}\} &= \delta \log \lambda(Y|\mathbf{x}) + \log S(Y|\mathbf{x}) \\ &= \delta \log \lambda(Y|\mathbf{x}) - \int \operatorname{ind}(Y \geq u)\, \lambda(u|\mathbf{x})\, du\end{aligned}$$

[see Kooperberg et al. (1995a)]. Observe that

$$E\Big(\int \operatorname{ind}(Y \geq u)\, \lambda(u|\mathbf{x})\, du\Big) = \int \lambda(u|\mathbf{x}) S_C(u|\mathbf{x}) S(u|\mathbf{x})\, du.$$

Thus the expected log-likelihood is given by

$$\begin{aligned}&E\Big(\delta \log \lambda(Y|\mathbf{x}) - \int \operatorname{ind}(Y \geq u)\, \lambda(u|\mathbf{x})\, du\Big) \\ &\qquad = \int S_C(t|\mathbf{x})[\log \lambda(t|\mathbf{x}) f(t|\mathbf{x}) - S(t|\mathbf{x})\lambda(t|\mathbf{x})]\, dt.\end{aligned}$$

Let $\mathbf{x}_i \in \mathbf{R}^M$ denote the vector of covariates for the ith individual, $1 \leq i \leq n$. Let $(T_1, C_1), \ldots, (T_n, C_n)$ be independent random vectors such that T_i and C_i are independent random variables having distribution functions $F(\cdot|\mathbf{x}_i)$ and $F_C(\cdot|\mathbf{x}_i)$, respectively. Set $Y_i = \min(T_i, C_i)$ and $\delta_i = \operatorname{ind}(T_i \leq C_i)$ for $1 \leq i \leq n$. Let G denote a linear space of functions on $\mathcal{T} \times \mathcal{X}$. The expected log-likelihood function $\Lambda(\cdot)$ is defined by

$$\Lambda(g) = \sum_i \int S_C(t|\mathbf{x}_i)[g(t|\mathbf{x}_i) f(t|\mathbf{x}_i) - S(t|\mathbf{x}_i) \exp g(t|\mathbf{x}_i)]\, dt, \qquad g \in G.$$

Observe that $\Lambda(\cdot)$ is maximized at $\phi(t|\mathbf{x}) = \log[f(t|\mathbf{x})/S(t|\mathbf{x})]$ which may or may not be in G. We define the best approximation to ϕ as a function in G that maximizes $\Lambda(\cdot)$ over G.

The first goal is to prove that $\Lambda(\cdot)$ has a maximum in G. Suppose the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of covariates take values in a compact interval $\mathcal{X} \subset \mathbf{R}^M$. Let $\mathcal{T}$ denote a compact interval of the form $[0, \tau]$ for some positive number τ. Without loss of generality, we assume that $\mathcal{T} = [0, 1]$ and $\mathcal{X} = [0, 1]^M$. The following conditions will be required to prove the existence of the best approximation in G.

Condition 1 *The density function $f(\cdot|\cdot)$ is bounded away from zero and infinity on $\mathcal{T} \times \mathcal{X}$. Moreover, the survival function $S(\cdot|\cdot)$ is bounded away from zero on $\mathcal{T} \times \mathcal{X}$.*

This condition implies that $S(1|\mathbf{x}) = P(T > 1 \mid \mathbf{x}) > 0$ on $\mathcal{X}$ and that $|\phi(\cdot|\cdot)|$ is bounded away from infinity on $\mathcal{T} \times \mathcal{X}$.

Condition 2 *$P(C \in \mathcal{T} \mid \mathbf{x}) = 1$ for $\mathbf{x} \in \mathcal{X}$ and $P(C = 1 \mid \mathbf{x})$ is bounded away from zero on $\mathcal{X}$.*

This condition implies that the censoring distribution $S_C(t|\mathbf{x})$ is bounded away from zero on $\mathcal{T} \times \mathcal{X}$. According to this condition, censoring automatically occurs at time 1 if failure or censoring does not occur before this time.

2.2 Polynomial splines and its tensor procducts

The best approximation will be chosen from the linear space G of tensor products of polynomial splines. Specifically, let $K = K_n$ be a positive integer and let I_k, $1 \leq k \leq K$, denote the subintervals of $[0, 1]$ defined by $I_k = [(k-1)/K, k/K)$ for $1 \leq k < K$ and $I_K = [1 - 1/K, 1]$. Let m and q be fixed integers such that $m \geq 0$ and $m > q \geq -1$. Let $\mathcal{S}$ denote the space of functions s on $[0, 1]$ such that

(i) the restriction of s to I_k is a polynomial of degree m (or less) for $1 \le k \le K$; and, if $q \ge 0$, then

(ii) s is q-times continuously differentiable on $[0, 1]$.

A function satisfying (i) is called a piecewise polynomial, and it is called a spline if it satisfies both (i) and (ii). Let B_j, $1 \le j \le J$, denote the usual basis of $\mathcal{S}$ consisting of B-splines [see de Boor (1978)]. Then $J = (m+1)K - (q+1)(K-1)$, so $K + m \le J \le (m+1)K$. Also, $B_j \ge 0$ on $[0,1]$, $B_j = 0$ on the complement of an interval of length $(m+1)/K$ for $1 \le j \le J$, and $\sum_j B_j = 1$ on $[0,1]$. Moreover, for $1 \le j \le J$, there are at most $2m+1$ values of $j' \in \{1, \dots, J\}$ such that $B_j B'_j$ is not identically zero on $[0,1]$. Set $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J) \in \mathbf{R}^J$ and let $|\boldsymbol{\beta}| = (\sum_j \beta_j^2)^{1/2}$ denote the Euclidean norm of $\boldsymbol{\beta}$. According to Theorem 4.2 of DeVore and Lorentz (1993), there is a positive constant M_0 such that

$$(2.1) \qquad M_0^{-1} J^{-1} |\boldsymbol{\beta}|^2 \le \int \Big| \sum_j \beta_j B_j \Big|^2 \le M_0 J^{-1} |\boldsymbol{\beta}|^2, \qquad \boldsymbol{\beta} \in \mathbf{R}^J.$$

Set $d = M + 1$. Let A denote the collection of ordered d-tuples $j = (j_0, j_1, \dots, j_M)$ with $j_0, j_1, \dots, j_M \in \{1, \dots, J\}$. Let G denote the tensor product space spanned by the basis functions on $\mathcal{T} \times \mathcal{X}$ of the form

$$B_j(t|\mathbf{x}) = B_{j_0}(t) \prod_{1 \le v \le M} B_{j_v}(x_v), \qquad \mathbf{x} = (x_1, \dots, x_M),$$

as j ranges over A. Then G has dimension $I := J^d$. Set $\mathbf{B} = \mathbf{B}(t|\mathbf{x}) = (B_j(t|\mathbf{x}))_{j \in A}$ and $\boldsymbol{\theta} = (\theta_j)_{j \in A} \in \Theta := \mathbf{R}^I$. Then each $g \in G$ can be written as $g(t|\mathbf{x}; \boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \mathbf{B}(t|\mathbf{x}) := \boldsymbol{\theta}^\top \mathbf{B}(t|\mathbf{x})$. (Here $\mathbf{A}^\top$ denotes the transpose of $\mathbf{A}$.)

Let $\mathbf{H}(\mathbf{x})$ denote the vector in $\mathbf{R}^{J^M}$ with entries $H_{j_1,\dots,j_M}(\mathbf{x}) = B_{j_1}(x_1) \cdots B_{j_M}(x_M)$, $j_1, \dots, j_M \in \{1, \dots, J\}$. Also, set

$$\boldsymbol{\beta} \cdot \mathbf{H}(\mathbf{x}) = \sum_{j_1=1}^{J} \cdots \sum_{j_M=1}^{J} \beta_{j_1,\dots,j_M} B_{j_1}(x_1) \cdots B_{j_M}(x_M),$$

where $\boldsymbol{\beta} = (\beta_{j_1,\ldots,j_M}) \in \mathbf{R}^{J^M}$. The following condition on the design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is required for the existence of the best spline approximation.

Condition 3 *There is a positive constant M_1 such that*

$$(2.2)\ M_1^{-1} n \int (\boldsymbol{\beta} \cdot \mathbf{H})^2 \leq \sum_i \left(\boldsymbol{\beta} \cdot \mathbf{H}(\mathbf{x}_i)\right)^2 \leq M_1 n \int (\boldsymbol{\beta} \cdot \mathbf{H})^2, \quad \boldsymbol{\beta} \in \mathbf{R}^{J^M},$$

and

$$(2.3) \qquad \sum_i H_{j_1,\ldots,j_M}(\mathbf{x}_i) \leq M_1 n J^{-M}, \qquad j_1, \ldots, j_M \in \{1, \ldots, J\}.$$

REMARK. It follows easily from Lemma 3.4 of Stone (1994) that if Condition 5 below holds and if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are replaced by independent random variables with a common density function that is bounded away from zero and infinity on $\mathcal{X}$ [as in Kooperberg et al. (1995b)], then (2.2) and (2.3) hold for sufficiently large n, except on an event whose probability tends to zero with n.

2.3 Spline approximation

Under Conditions 1–3, there exists an essentially uniquely determined function $\phi^* \in G$ such that $\Lambda(\phi^*) = \max_{g \in G} \Lambda(g)$. Moreover, if $\phi \in G$, then $\phi^* = \phi$ almost everywhere [see Kooperberg et al. (1995b)]. Let $\boldsymbol{\theta}^*$ denote the vector of parameters that is associated with ϕ^*, so that $\phi^*(t|\mathbf{x}) = \boldsymbol{\theta}^* \cdot \mathbf{B}(t|\mathbf{x})$, and set $\lambda^*(t|\mathbf{x}) = \exp \phi^*(t|\mathbf{x})$. Also, set $H^*(t|\mathbf{x}) = \int_0^t \lambda^*(u|\mathbf{x})\, du$, $S^*(t|\mathbf{x}) = \exp(-H^*(t|\mathbf{x}))$ and $f^*(t|\mathbf{x}) = \lambda^*(t|\mathbf{x})\, S^*(t|\mathbf{x})$. These functions are referred to as spline approximations.

The errors resulting from spline approximation will be quantified in terms of a smoothness condition that will now be described. Let $0 < \beta \leq 1$. A function g on $\mathcal{T} \times \mathcal{X}$ is said to satisfy a Hölder condition with exponent β if there is a positive number γ such that $|g(\mathbf{z}) - g(\mathbf{z}_0)| \leq \gamma |\mathbf{z} - \mathbf{z}_0|^\beta$ for $\mathbf{z}, \mathbf{z}_0 \in \mathcal{T} \times \mathcal{X}$. Let m be a nonnegative integer and set $p = m + \beta$. A function g on $\mathcal{T} \times \mathcal{X}$ is said to be p-smooth if it is m times continuously differentiable on $\mathcal{T} \times \mathcal{X}$ and $g^{(m)}$ satisfies a Hölder condition with exponent β. The following smoothness condition will be used to describe the errors resulting from spline approximation.

Condition 4 *ϕ is a p-smooth function with $p > d/2$.*

We do not assume that ϕ is exactly equal to a spline, but we still can make use of spline approximation. In order for this method to be accurate, we need the error of approximation to tend to zero as the sample size n tends to infinity; for this, it is necessary that the dimension I of the approximation space G tend to infinity. To control the error of estimation we need this dimension to increase more slowly than $n^{1/2}$.

Condition 5 *$I = I_n \to \infty$ and $I^2 = o(n^{1-\epsilon})$ for some $\epsilon > 0$.*

For a real-valued function h on $\mathcal{T} \times \mathcal{X}$, set $\|h\|_2 = [\int_{\mathcal{T}\times\mathcal{X}} |h(t|\mathbf{x})|^2]^{1/2}$ and $\|h\|_\infty = \sup_{\mathcal{T}\times\mathcal{X}} |h(\cdot|\cdot)|$. Also, set $\rho = \inf_{g\in G} \|g - \phi\|_\infty$. Under the first part of Condition 5, $\rho = o(1)$ [see (2) on page 167 of de Boor (1978)]. Under Conditions 4 and 5, $\rho = O(J^{-p})$ [see Theorem XII.1 of de Boor (1978)]. Our first result gives the error bounds for the spline approximations.

Theorem 1 *Under Conditions 1–5,*

$$\|\phi^* - \phi\|_\infty = O(\rho), \tag{2.4}$$

$$\|\lambda^* - \lambda\|_\infty = O(\rho), \tag{2.5}$$

$$\|H^* - H\|_\infty = O(\rho), \tag{2.6}$$

$$\|S^* - S\|_\infty = O(\rho), \tag{2.7}$$

$$\|f^* - f\|_\infty = O(\rho). \tag{2.8}$$

The proof of Theorem 1 is given in Section 3.

2.4 Maximum likelihood estimation

The likelihood corresponding to the survival data $(Y_1, \delta_1, \mathbf{x}_1), \ldots, (Y_n, \delta_n, \mathbf{x}_n)$ equals

$$\prod_i \left\{ [\lambda(Y_i|\mathbf{x}_i)]^{\delta_i} S(Y_i|\mathbf{x}_i) \right\},$$

and the log-likelihood is given by

$$\ell(g) = \sum_i \left(\delta_i g(Y_i|\mathbf{x}_i;\boldsymbol{\theta}) - \int_0^{Y_i} \exp g(u|\mathbf{x}_i;\boldsymbol{\theta})\, du \right), \qquad g = g(\cdot|\cdot;\boldsymbol{\theta}) \in G.$$

Under Conditions 1–5, the log-likelihood function ℓ is strictly concave and hence there exists a unique maximum likelihood estimate $\widehat{g} = \widehat{\boldsymbol{\theta}} \cdot \mathbf{B} \in G$ [see Kooperberg et al. (1995b)], so that $\widehat{\boldsymbol{\theta}} = \max_\Theta \ell(\boldsymbol{\theta})$. The maximum likelihood estimates of the log-hazard, hazard, cumulative hazard, survival and density functions are given, respectively, by $\widehat{\phi}(t|\mathbf{x}) = \widehat{\boldsymbol{\theta}} \cdot \mathbf{B}(t|\mathbf{x})$, $\widehat{\lambda}(t|\mathbf{x}) = \exp\, \widehat{\phi}(t|\mathbf{x})$, $\widehat{H}(t|\mathbf{x}) = \int_0^t \widehat{\lambda}(u|\mathbf{x})\, du$, $\widehat{S}(t|\mathbf{x}) = \exp(-\widehat{H}(t|\mathbf{x}))$ and $\widehat{f}(t|\mathbf{x}) = \widehat{\lambda}(t|\mathbf{x})\widehat{S}(t|\mathbf{x})$ for $0 \le t \le 1$. These estimates are referred to as HARE estimates. The next result bounds the L_2 and L_∞ norms of the error of the various HARE estimates.

Theorem 2 *Under Conditions* 1–5,

$$|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*| = O_p(I/\sqrt{n}), \tag{2.9}$$

$$\|\widehat{\phi} - \phi^*\|_2 = O_p(\sqrt{n^{-1}I}), \tag{2.10}$$

$$\|\widehat{\lambda} - \lambda^*\|_2 = O_p(\sqrt{n^{-1}I}), \tag{2.11}$$

$$\max_{j \in A} |\widehat{\theta}_j - \theta_j^*|^2 = O_p(n^{-1} I \log I), \tag{2.12}$$

$$\max_{t \in \mathcal{T}, \mathbf{x} \in \mathcal{X}} |\widehat{\phi}(t|\mathbf{x}) - \phi^*(t|\mathbf{x})| = O_p\left(\sqrt{n^{-1} I \log I}\right), \tag{2.13}$$

$$\max_{t \in \mathcal{T}, \mathbf{x} \in \mathcal{X}} |\widehat{\lambda}(t|\mathbf{x}) - \lambda^*(t|\mathbf{x})| = O_p\left(\sqrt{n^{-1} I \log I}\right), \tag{2.14}$$

$$\max_{t \in \mathcal{T}} |\widehat{H}(t|\mathbf{x}) - H^*(t|\mathbf{x})| = O_p\left(\sqrt{n^{-1} J^M}\right), \qquad \mathbf{x} \in \mathcal{X} \tag{2.15}$$

$$\max_{t \in \mathcal{T}} |\widehat{S}(t|\mathbf{x}) - S^*(t|\mathbf{x})| = O_p\left(\sqrt{n^{-1} J^M}\right), \qquad \mathbf{x} \in \mathcal{X} \tag{2.16}$$

$$
\begin{aligned}
(2.17)\qquad & \|\widehat{f}-f^*\|_2 &=& \; O_p(\sqrt{n^{-1}I}), \\
(2.18)\qquad & \max_{t\in\mathcal{T},\mathbf{x}\in\mathcal{X}} |\widehat{f}(t|\mathbf{x})-f^*(t|\mathbf{x})| &=& \; O_p\Big(\sqrt{n^{-1}I\log I}\Big).
\end{aligned}
$$

The proof of the above theorem is given in Section 4. Let $J \sim n^{1/(2p+d)}$ in Theorems 1 and 2. (Here $a_n \sim b_n$ means that a_n/b_n is bounded away from zero and infinity.) Then $\rho = O(J^{-p})$. We conclude from these theorems that

$$
\|\widehat{\phi}-\phi\|_2 = O_p(n^{-p/(2p+d)}), \qquad \|\widehat{\phi}-\phi\|_\infty = O_p\Big(\{(\log\, n)/n\}^{-p/(2p+d)}\Big);
$$

$$
\|\widehat{\lambda}-\lambda\|_2 = O_p(n^{-p/(2p+d)}), \qquad \|\widehat{\lambda}-\lambda\|_\infty = O_p\Big(\{(\log\, n)/n\}^{-p/(2p+d)}\Big);
$$

$$
\|\widehat{f}-f\|_2 = O_p(n^{-p/(2p+d)}), \qquad \|\widehat{f}-f\|_\infty = O_p\Big(\{(\log\, n)/n\}^{-p/(2p+d)}\Big).
$$

Under suitable conditions, these are the well known optimal rates of convergence for nonparametric function estimation; see Stone (1980, 1982, 1983).

It follows from (2.6), (2.7), (2.15) and (2.16) that if $J \sim n^{1/(2p+M)}$ with $p > (d+1)/2$, then $\max_t |\widehat{H}(t|\mathbf{x}) - H(t|\mathbf{x})| = O_p(n^{-p/(2p+M)})$ and $\max_t |\widehat{S}(t|\mathbf{x}) - S(t|\mathbf{x})| = O_p(n^{-p/(2p+M)})$. In particular, these rates are $n^{-1/2}$ in the absence of covariates ($M=0$). These are well known optimal rates for estimation of the cumulative hazard and survival functions in both parametric and nonparametric settings [see Breslow and Crowley (1974), and Andersen et al. (1993)].

2.5 Asymptotic distributions of HARE estimates

Let $\mathbf{I} = \mathbf{I}(\boldsymbol{\theta})$ denote the $I \times I$ information matrix, which has entries

$$
-E\Big(\frac{\partial^2 \ell(g)}{\partial\theta_j \partial\theta_k}\Big) = \sum_i E\Big(\int_0^{Y_i} B_j(u|\mathbf{x}_i)B_k(u|\mathbf{x}_i)\exp\, g(u|\mathbf{x}_i;\boldsymbol{\theta})\, du\Big)
$$

as j and k range over A. Let ω denote a real-valued parameter depending on ϕ^*, so that $\omega = \Gamma(\boldsymbol{\theta}^*)$ for some function $\Gamma(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. The maximum likelihood estimate of ω is given by $\widehat{\omega} = \Gamma(\widehat{\boldsymbol{\theta}})$. Suppose Γ is continuously differentiable on Θ. Let $\nabla\Gamma(\boldsymbol{\theta})$ denote the gradient of Γ at $\boldsymbol{\theta}$, which is the I-dimensional vector whose jth entry is $\partial\Gamma(\boldsymbol{\theta})/\partial\theta_j$. The asymptotic standard deviation (ASD) and standard error (SE) of $\widehat{\omega}$ are defined by

$$\mathrm{ASD}(\widehat{\omega}) = \sqrt{\nabla\Gamma(\boldsymbol{\theta}^*)^\top[\mathbf{I}(\boldsymbol{\theta}^*)]^{-1}\nabla\Gamma(\boldsymbol{\theta}^*)}$$

and

$$\mathrm{SE}(\widehat{\omega}) = \sqrt{\nabla\Gamma(\widehat{\boldsymbol{\theta}})^\top[\mathbf{I}(\widehat{\boldsymbol{\theta}})]^{-1}\nabla\Gamma(\widehat{\boldsymbol{\theta}})}.$$

When ω depends on t and $\mathbf{x}$, we write $\Gamma(\boldsymbol{\theta})$ as $\Gamma(\boldsymbol{\theta}; t, \mathbf{x})$. Some important cases are given in the following table:

ω	$\Gamma(\boldsymbol{\theta}; t, \mathbf{x})$	$\nabla\Gamma(\boldsymbol{\theta}; t, \mathbf{x})$
$\phi(t\|\mathbf{x})$	$\boldsymbol{\theta}\cdot\mathbf{B}(t\|\mathbf{x})$	$\mathbf{B}(t\|\mathbf{x})$
$\lambda(t\|\mathbf{x})$	$\exp(\boldsymbol{\theta}\cdot\mathbf{B}(t\|\mathbf{x}))$	$\mathbf{B}(t\|\mathbf{x})\exp(\boldsymbol{\theta}\cdot\mathbf{B}(t\|\mathbf{x}))$
$H(t\|\mathbf{x})$	$\int_0^t \exp(\boldsymbol{\theta}\cdot\mathbf{B})$	$\int_0^t \mathbf{B}\exp(\boldsymbol{\theta}\cdot\mathbf{B})$
$S(t\|\mathbf{x})$	$\exp[-\int_0^t \exp(\boldsymbol{\theta}\cdot\mathbf{B})]$	$-\Gamma(\boldsymbol{\theta}; t, \mathbf{x})\int_0^t \mathbf{B}\exp(\boldsymbol{\theta}\cdot\mathbf{B})$

Thus, for example,

$$\mathrm{ASD}(\widehat{\phi}(t|\mathbf{x})) = \sqrt{[\mathbf{B}(t|\mathbf{x})]^\top[\mathbf{I}(\boldsymbol{\theta}^*)]^{-1}\mathbf{B}(t|\mathbf{x})}$$

and

$$\mathrm{SE}(\widehat{\phi}(t|\mathbf{x})) = \sqrt{[\mathbf{B}(t|\mathbf{x})]^\top[\mathbf{I}(\widehat{\boldsymbol{\theta}})]^{-1}\mathbf{B}(t|\mathbf{x})}.$$

To obtain the asymptotic standard deviation and the standard error of the density estimate, set $\Gamma(\boldsymbol{\theta}; t, \mathbf{x}) = \exp(\boldsymbol{\theta}\cdot\mathbf{B}(t|\mathbf{x}))\exp[-\int_0^t \exp(\boldsymbol{\theta}\cdot\mathbf{B})]$ for $0 \le t \le 1$. Then $\nabla\Gamma(\boldsymbol{\theta}; t, \mathbf{x}) = \Gamma(\boldsymbol{\theta}; t, \mathbf{x})\,[\mathbf{B}(t|\mathbf{x}) - \int_0^t \mathbf{B}\exp(\boldsymbol{\theta}\cdot\mathbf{B})]$. Thus, $\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x}) = f^*(t|\mathbf{x})$ and $\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x}) = f^*(t|\mathbf{x})\,[\mathbf{B}(t|\mathbf{x}) - \int_0^t \mathbf{B}\lambda^*]$. In fact, it follows from the basic properties of B-splines that

$$\mathrm{ASD}(\widehat{f}(t|\mathbf{x})) \simeq f^*(t|\mathbf{x})\sqrt{[\mathbf{B}(t|\mathbf{x})]^\top[\mathbf{I}(\boldsymbol{\theta}^*)]^{-1}\mathbf{B}(t|\mathbf{x})}$$

and

$$\mathrm{SE}(\widehat{f}(t|\mathbf{x})) \simeq \widehat{f}(t|\mathbf{x})\sqrt{[\mathbf{B}(t|\mathbf{x})]^\top[\mathbf{I}(\widehat{\boldsymbol{\theta}})]^{-1}\mathbf{B}(t|\mathbf{x})}, \qquad 0 \le t \le 1.$$

(Here $a_n \simeq b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.)

The asymptotic distributions of HARE estimates are described in the following result. Here $Z_n \xrightarrow{d} N(0,1)$ means that the distribution of Z_n converges to the standard normal distribution as $n \to \infty$.

Theorem 3 *Under Conditions* 1–5, *for* $t \in \mathcal{T}$ *and* $\mathbf{x} \in \mathcal{X}$,

$$\frac{\widehat{\phi}(t|\mathbf{x}) - \phi^*(t|\mathbf{x})}{\mathrm{ASD}(\widehat{\phi}(t|\mathbf{x}))} \xrightarrow{d} N(0,1), \quad \frac{\mathrm{SE}(\widehat{\phi}(t|\mathbf{x}))}{\mathrm{ASD}(\widehat{\phi}(t|\mathbf{x}))} = 1 + o_p(1), \tag{2.19}$$

$$\frac{\widehat{\lambda}(t|\mathbf{x}) - \lambda^*(t|\mathbf{x})}{\mathrm{ASD}(\widehat{\lambda}(t|\mathbf{x}))} \xrightarrow{d} N(0,1), \quad \frac{\mathrm{SE}(\widehat{\lambda}(t|\mathbf{x}))}{\mathrm{ASD}(\widehat{\lambda}(t|\mathbf{x}))} = 1 + o_p(1), \tag{2.20}$$

$$\frac{\widehat{H}(t|\mathbf{x}) - H^*(t|\mathbf{x})}{\mathrm{ASD}(\widehat{H}(t|\mathbf{x}))} \xrightarrow{d} N(0,1), \quad \frac{\mathrm{SE}(\widehat{H}(t|\mathbf{x}))}{\mathrm{ASD}(\widehat{H}(t|\mathbf{x}))} = 1 + o_p(1), \quad t > 0, \tag{2.21}$$

$$\frac{\widehat{S}(t|\mathbf{x}) - S^*(t|\mathbf{x})}{\mathrm{ASD}(\widehat{S}(t|\mathbf{x}))} \xrightarrow{d} N(0,1), \quad \frac{\mathrm{SE}(\widehat{S}(t|\mathbf{x}))}{\mathrm{ASD}(\widehat{S}(t|\mathbf{x}))} = 1 + o_p(1), \quad t > 0, \tag{2.22}$$

$$\frac{\widehat{f}(t|\mathbf{x}) - f^*(t|\mathbf{x})}{\mathrm{ASD}(\widehat{f}(t|\mathbf{x}))} \xrightarrow{d} N(0,1), \quad \frac{\mathrm{SE}(\widehat{f}(t|\mathbf{x}))}{\mathrm{ASD}(\widehat{f}(t|\mathbf{x}))} = 1 + o_p(1). \tag{2.23}$$

The proof of Theorem 3 is given in Section 5. Confidence intervals can be constructed using Theorem 3 in an obvious manner. Suppose $\omega = \Gamma(\boldsymbol{\theta}^*)$ is a parameter of interest. Then $\widehat{\omega} \pm z_{1-\alpha}\,\mathrm{SE}(\widehat{\omega})$ is an asymptotic $100(1-\alpha)\%$ confidence interval for ω. Here $\Phi(z_{1-\alpha}) = 1 - \alpha$ with Φ being the standard normal distribution function. In particular, the $100(1-\alpha)\%$ confidence interval for $\phi^*(t|\mathbf{x})$ is given by

$$\widehat{\phi}(t|\mathbf{x}) \pm z_{1-\alpha}\sqrt{[\mathbf{B}(t|\mathbf{x})]^\top[\mathbf{I}(\widehat{\boldsymbol{\theta}})]^{-1}\mathbf{B}(t|\mathbf{x})}.$$

Note that this will be the $100(1-\alpha)\%$ confidence interval for $\phi(t|\mathbf{x})$ only when the error of approximation (quantified in terms of ρ as described in Theorem 1) goes to zero faster than the asymptotic standard deviation of $\widehat{\phi}$. This can be achieved by choosing J such that $Jn^{-1/(2p+d)} \gg 1$. If ρ goes to zero at a rate slower or equivalent to that of $\mathrm{ASD}(\widehat{\phi})$, then the above interval is the $100(1-\alpha)\%$ confidence interval for the best approximation to the log-hazard function. Similar remarks also apply to the confidence intervals for the cumulative hazard, survival and density functions.

The above asymptotic results are analogous to those given in Stone (1991) for logspline response model estimation based on noncensored data. The proofs given here are mostly similar to the proofs in Stone (1991). However, Lemmas 14 and 22 of Stone (1991) were established by using well-known properties of exponential families. The corresponding results in this paper, Lemmas 3 and 8, are the key to obtaining the L_∞ rates of convergence and asymptotic normality for the various HARE estimates. We prove Lemma 3 by using a result motivated by the elementary theory of differential equations (see Lemma 2) and Lemma 8 by using the martingale structure of the log-likelihood function.

3 Proof of Theorem 1

We assume that Conditions 1–4 hold throughout this section. The proof of (2.4), which follows from an argument similar to that given in Stone (1989) and based on de Boor (1976), is given in the appendix for completeness. Equation (2.5) follows from (2.4) and the following:

$$\lambda^* - \lambda = \exp\phi^* - \exp\phi = (\phi^* - \phi)\int_0^1 \exp(\phi + u(\phi^* - \phi))\,du.$$

Equation (2.6) follows easily from (2.5).

The proof of (2.7) follows from (2.6) and

$$S^* - S = \exp(-H^*) - \exp(-H) = -(H^* - H)\int_0^1 \exp(-H - u(H^* - H))\,du.$$

Observe that $f^* - f = \lambda^*(S^* - S) + (\lambda^* - \lambda)S$. We conclude from (2.5) and (2.7) that (2.8) holds.

4 Proof of Theorem 2

We assume Conditions 1–5 hold throughout this section. Recall that the log-likelihood function is given by

$$\ell(g) = \sum_i \delta_i g(Y_i|\mathbf{x}_i;\boldsymbol{\theta}) - \sum_i \int_0^{Y_i} \exp g(u|\mathbf{x}_i;\boldsymbol{\theta})\,du.$$

Let

$$\mathbf{S}(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}\ell(g)$$

denote the score at $\boldsymbol{\theta}$, which is the I-dimensional vector having entries

$$\frac{\partial \ell(g)}{\partial\theta_j} = \sum_i \delta_i B_j(Y_i|\mathbf{x}_i) - \sum_i \int_0^{Y_i} B_j(u|\mathbf{x}_i) \exp g(u|\mathbf{x}_i;\boldsymbol{\theta})\,du.$$

Let

$$\frac{\partial^2 \ell(g)}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^\top}$$

denote the Hessian of $\ell(g)$, which is the $I \times I$ matrix having entries

$$\frac{\partial^2 \ell(g)}{\partial\theta_j\partial\theta_k} = -\sum_i \int_0^{Y_i} B_j(u|\mathbf{x}_i)B_k(u|\mathbf{x}_i) \exp g(u|\mathbf{x}_i;\boldsymbol{\theta})\,du.$$

[Recall that $j = (j_0, j_1, \ldots, j_d)$ and $k = (k_0, k_1, \ldots, k_d)$.]

Set $\mathbf{S}^* = \mathbf{S}(\boldsymbol{\theta}^*)$. The maximum likelihood equation $\mathbf{S}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$ can be written as

$$\int_0^1 \frac{d}{du}\mathbf{S}(\boldsymbol{\theta}^* + u(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*))\,du = -\mathbf{S}^*.$$

This can further be written as $\mathbf{D}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = -\mathbf{S}^*$, where $\mathbf{D}$ is the $I \times I$ matrix given by

$$\mathbf{D} = \int_0^1 \frac{\partial^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^\top}\ell(\boldsymbol{\theta}^* + u(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*))\,du. \tag{4.1}$$

Proofs of (2.9) and (2.10)

It follows from the maximum likelihood equation that

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^\top \mathbf{D}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = -(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^\top \mathbf{S}^*. \tag{4.2}$$

According to (3.5) and (3.6) of Kooperberg et al. (1995b), there is positive constant M_1 such that

$$|\mathbf{S}^*|^2 = O_p(n) \tag{4.3}$$

and

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^\top \mathbf{D}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \le -M_1 n I^{-1} |\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*|^2 \tag{4.4}$$

except on an event whose probability tends to zero with n. Since

$$|(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^\top \mathbf{S}^*| \le |\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*|\,|\mathbf{S}^*|,$$

it follows from (4.2)–(4.4) that $|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*|^2 = O_p(I^2/n)$ and hence that

$$\|\widehat{\phi} - \phi^*\|^2 = O_p(I/n).$$

This completes the proofs of (2.9) and (2.10).

Proof of (2.11)

According to (2.10), Lemma 2 of Kooperberg et al. (1995b) and Condition 5, $\|\widehat{\phi} - \phi^*\|_\infty = O_p(I/\sqrt{n}) = o_p(1)$. The desired result follows from (2.4), (2.10) and

$$\widehat{\lambda} - \lambda^* = \exp\widehat{\phi} - \exp\phi^* = (\widehat{\phi} - \phi^*) \int_0^1 \exp(\phi^* + u(\widehat{\phi} - \phi^*))\, du. \tag{4.5}$$

The proof of (2.12) requires a sequence of lemmas. Set $\lambda_{\min}(\boldsymbol{\theta}) = \min[\exp g(\cdot|\cdot;\boldsymbol{\theta})]$ and $\lambda_{\max}(\boldsymbol{\theta}) = \max[\exp g(\cdot|\cdot;\boldsymbol{\theta})]$.

Lemma 1 *There is a positive constant M_2 such that*

$$M_2^{-1}\lambda_{\min}(\boldsymbol{\theta})\, nI^{-1}|\boldsymbol{\tau}|^2 \le \boldsymbol{\tau}^\top \mathbf{I}(\boldsymbol{\theta})\boldsymbol{\tau} \le M_2\lambda_{\max}(\boldsymbol{\theta})\, nI^{-1}|\boldsymbol{\tau}|^2 \tag{4.6}$$

for $\boldsymbol{\theta}, \boldsymbol{\tau} \in \Theta$ and $n \ge 1$. Moreover,

$$[M_2\lambda_{\max}(\boldsymbol{\theta})]^{-1} n^{-1} I|\boldsymbol{\tau}|^2 \le \boldsymbol{\tau}^\top [\mathbf{I}(\boldsymbol{\theta})]^{-1}\boldsymbol{\tau} \le M_2[\lambda_{\min}(\boldsymbol{\theta})]^{-1} n^{-1} I|\boldsymbol{\tau}|^2 \tag{4.7}$$

and

$$[M_2\lambda_{\max}(\boldsymbol{\theta})]^{-1} n^{-1} I|\boldsymbol{\tau}| \le |[\mathbf{I}(\boldsymbol{\theta})]^{-1}\boldsymbol{\tau}| \le M_2[\lambda_{\min}(\boldsymbol{\theta})]^{-1} n^{-1} I|\boldsymbol{\tau}| \tag{4.8}$$

for $n \ge 1$ and $\boldsymbol{\theta}, \boldsymbol{\tau} \in \Theta$ such that $\lambda_{\min}(\boldsymbol{\theta}) > 0$.

Proof. According to (3.9) of Kooperberg et al. (1995b),

$$\boldsymbol{\tau}^{\mathsf{T}}\mathbf{I}(\boldsymbol{\theta})\boldsymbol{\tau} = \sum_i \int_0^1 g^2(u|\mathbf{x}_i;\boldsymbol{\tau})S(u|\mathbf{x}_i)S_C(u|\mathbf{x}_i)\exp g(u|\mathbf{x}_i;\boldsymbol{\theta})\,du.$$

We conclude from (2.1), Condition 1 and (2.2) that (4.6) holds.

If $\lambda_{\min}(\boldsymbol{\theta}) > 0$, then it follows from (4.6) that there is a nonsingular symmetric matrix $\mathbf{R}(\boldsymbol{\theta})$ such that $\mathbf{I}(\boldsymbol{\theta}) = \mathbf{R}(\boldsymbol{\theta})\mathbf{R}(\boldsymbol{\theta})$. Also,

$$[M_2\lambda_{\max}(\boldsymbol{\theta})]^{-1}n^{-1}I \le \frac{|\boldsymbol{\tau}|^2}{\boldsymbol{\tau}^{\mathsf{T}}\mathbf{I}(\boldsymbol{\theta})\boldsymbol{\tau}} \le M_2[\lambda_{\min}(\boldsymbol{\theta})]^{-1}n^{-1}I, \quad \boldsymbol{\tau}\in\Theta.$$

Replacing $\boldsymbol{\tau}$ by $[\mathbf{R}(\boldsymbol{\theta})]^{-1}\boldsymbol{\tau}$, we conclude that (4.7) is valid.

Similarly, it follows from (4.7) applied to $\boldsymbol{\tau}$ and $[\mathbf{R}(\boldsymbol{\theta})]^{-1}\boldsymbol{\tau}$ that (4.8) is valid.

Lemma 2 *Let $\psi(\cdot)$ and $s(\cdot)$ denote piecewise smooth functions on* $[0,1]$. *Set*

$$w(y) = s(y) - \int_0^y s(u)\psi(u)\,du + \int_0^1 s(u)\psi(u)\,du = s(y) + \int_y^1 s(u)\psi(u)\,du.$$

Then

$$\int_0^1 s^2(y)\,dy = O\Big(\int_0^1 w^2(y)\,dy\Big).$$

Proof. We have

$$\begin{aligned}
ds(y) - s(y)\psi(y) &= dw(y),\\
d\Big(s(y)\exp\int_y^1\psi(u)\,du\Big) &= \Big(\exp\int_y^1\psi(u)\,du\Big)dw(y),\\
s(1) - s(y)\exp\int_y^1\psi(u)\,du &= \int_y^1\Big(\exp\int_u^1\psi(t)\,dt\Big)dw(u),\\
w(1) - s(y)\exp\int_y^1\psi(u)\,du &= \int_y^1\Big(\exp\int_u^1\psi(t)\,dt\Big)dw(u).
\end{aligned}$$

[Since $w(1) = s(1)$.] Thus

$$s(y)\exp\int_y^1\psi(u)\,du = w(1) - \int_y^1\Big(\exp\int_u^1\psi(t)\,dt\Big)\,dw(u)$$

$$
\begin{aligned}
&= w(1) - \Big[w(u) \exp \int_u^1 \psi(t)\, dt\Big]_y^1 \\
&\qquad - \int_y^1 w(u)\psi(u) \Big(\exp \int_u^1 \psi(t)\, dt\Big)\, du \\
&= w(y)\Big(\exp \int_y^1 \psi(u)\, du\Big) \\
&\qquad - \int_y^1 w(u)\psi(u)\Big(\exp \int_u^1 \psi(t)\, dt\Big)\, du.
\end{aligned}
$$

Set $\Psi(u,y) = \psi(u) \exp \int_u^y \psi(t)\, dt$. Then

$$
\begin{aligned}
s(y) &= w(y) - \exp\Big(-\int_y^1 \psi(u)\, du\Big) \int_y^1 w(u)\psi(u)\Big(\exp \int_u^1 \psi(t)\, dt\Big)\, du \\
&= w(y) - \int_y^1 w(u)\Psi(u,y)\, du.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\int_0^1 s^2(y)\, dy &\le 2\int_0^1 w^2(y)\, dy + 2\int_0^1 \Big(\int_y^1 w(u)\Psi(u,y)\, du\Big)^2 dy \\
&\le 2\int_0^1 w^2(y)\, dy + 2\int_0^1 w^2(u)\, du \int_0^1\int_0^1 \Psi^2(u,y)\, du\, dy \\
&= O\Big(\int_0^1 w^2(y)\, dy\Big).
\end{aligned}
$$

This completes the proof of Lemma 2.

Let $\mathbf{G}^*(y,\delta,\mathbf{x}) = \mathbf{G}(y,\delta,\mathbf{x};\boldsymbol{\theta}^*) = [G_j(y,\delta,\mathbf{x};\boldsymbol{\theta}^*)]$ denote the I-dimensional vectors with the j-th entry given by

$$
G_j(y,\delta,\mathbf{x};\boldsymbol{\theta}^*) = \Big(\delta B_{j_0}(y) - \int_0^y B_{j_0}(u) \exp g(u|\mathbf{x};\boldsymbol{\theta}^*)\, du\Big) \prod_{1\le v\le M} B_{j_v}(x_v),
$$

where $j = (j_0, j_1, \ldots, j_M)$ and $\mathbf{x} = (x_1, \ldots, x_M)$. It follows from (2.5) and the basic properties of B-splines that

$$
\max_{1\le j_0\le J} \sup_{y,\mathbf{x}} \Big| \int_0^y B_{j_0}(u) \exp g(u|\mathbf{x};\boldsymbol{\theta}^*)\, du\Big| = O(J^{-1}) \tag{4.9}
$$

and hence that

$$
|\mathbf{G}^*(y,\delta,\mathbf{x})| = O(1) \quad \text{uniformly in } (y,\mathbf{x}) \in \mathcal{T} \times \mathcal{X}. \tag{4.10}
$$

Note that $\mathbf{S}^* = \sum_i \mathbf{G}^*(Y_i, \delta_i, \mathbf{x}_i)$ and $E(\mathbf{S}^*) = \mathbf{0}$. Let $\mathrm{VC}(\mathbf{S}^*)$ denote the variance-covariance matrix of $\mathbf{S}^*$.

Lemma 3 *There is a positive constant M_3 such that*

$$M_3^{-1} n I^{-1} |\boldsymbol{\tau}|^2 \le \boldsymbol{\tau}^\top \mathrm{VC}(\mathbf{S}^*) \boldsymbol{\tau} \le M_3 n I^{-1} |\boldsymbol{\tau}|^2, \quad \boldsymbol{\tau} \in \Theta, \quad n \ge 1.$$

Proof. Note that

$$\begin{aligned} \boldsymbol{\tau}^\top \mathrm{VC}(\mathbf{S}^*) \boldsymbol{\tau} &= \mathrm{var}(\boldsymbol{\tau}^\top \mathbf{S}^*) \\ &= \sum_i \mathrm{var}\Big(\delta_i g(Y_i|\mathbf{x}_i; \boldsymbol{\tau}) - \int_0^{Y_i} g(u|\mathbf{x}_i; \boldsymbol{\tau}) \exp g(u|\mathbf{x}_i; \boldsymbol{\theta}^*)\, du\Big). \end{aligned}$$

By (2.1) and (2.2),

$$\sum_i E[g^2(Y_i|\mathbf{x}_i; \boldsymbol{\tau})] = O\Big(I^{-1}|\boldsymbol{\tau}|^2\Big) \tag{4.11}$$

and

$$\sum_i E\Big[\Big(\int_0^{Y_i} g(u|\mathbf{x}_i; \boldsymbol{\tau}) \exp g(u|\mathbf{x}_i; \boldsymbol{\theta}^*)\, du\Big)^2\Big] = O\Big(I^{-1}|\boldsymbol{\tau}|^2\Big). \tag{4.12}$$

It follows from (4.11) and (4.12) that the upper bound holds.

Set $s(y|\mathbf{x}) = g(y|\mathbf{x}; \boldsymbol{\tau})$, $\psi(y|\mathbf{x}) = \exp g(y|\mathbf{x}; \boldsymbol{\theta}^*)$ and $\mu = E[s(Y|\mathbf{x}) - \int_0^Y s(u|\mathbf{x})\psi(u|\mathbf{x})]\, du$. Since censoring automatically occurs at time $t = 1$ (Condition 2), we have

$$\begin{aligned} &\mathrm{var}\Big(\delta_i g(Y|\mathbf{x}; \boldsymbol{\tau}) - \int_0^Y g(u|\mathbf{x}; \boldsymbol{\tau}) \exp g(u|\mathbf{x}; \boldsymbol{\theta}^*)\, du\Big) \\ &= \mathrm{var}\Big(\delta s(Y|\mathbf{x}) - \int_0^Y s(u|\mathbf{x})\psi(u|\mathbf{x})\, du\Big) \\ &= E\Big[\Big(\delta s(Y|\mathbf{x}) - \int_0^Y s(u|\mathbf{x})\psi(u|\mathbf{x})\, du - \mu\Big)^2; Y < 1\Big] \\ &\quad + E\Big[\Big(-\int_0^Y s(u|\mathbf{x})\psi(u|\mathbf{x})\, du - \mu\Big)^2; Y \ge 1\Big] \\ &\ge \iint_{t \le c} \Big[s(t) - \int_0^t s(u|\mathbf{x})\psi(u|\mathbf{x})\, du - \mu\Big]^2 dS_C(c|\mathbf{x})\, dS(t|\mathbf{x}) \end{aligned} \tag{4.13}$$

$$
\begin{aligned}
&+P(Y=1)\Big[-\int_0^1 s(u|\mathbf{x})\psi(u|\mathbf{x})\,du-\mu\Big]^2\\
\ge \min_{0\le t\le 1} f(t|\mathbf{x})&\int_0^1 S_C(t|\mathbf{x})\Big[s(t)-\int_0^t s(u|\mathbf{x})\psi(u|\mathbf{x})\,du-\mu\Big]^2\,dt\\
&+P(Y=1)\int_0^1\Big[-\int_0^1 s(u|\mathbf{x})\psi(u|\mathbf{x})\,du-\mu\Big]^2\,dt.
\end{aligned}
$$

By Conditions 1 and 2, $P(Y=1)=S_C(1|\mathbf{x})>0$ and $S_C(t|\mathbf{x})\ge S_C(1|\mathbf{x})>0$ for $0\le t\le 1$. Set $c_1=\min\{\min_{0\le t\le 1} f(t|\mathbf{x})S_C(1|\mathbf{x}), P(Y=1)\}$. Then $c_1>0$. Hence by (4.13), Lemma 2 and Condition 1, there is a positive constant c_2 such that

$$
\begin{aligned}
&\operatorname{var}\Big(\delta_i g(Y|\mathbf{x};\boldsymbol{\tau})-\int_0^Y g(u|\mathbf{x};\boldsymbol{\tau})\exp g(u|\mathbf{x};\boldsymbol{\theta}^*)\,du\Big)\\
&\ge \frac{c_1}{2}\int_0^1\Big[s(t|\mathbf{x})-\int_0^t s(u|\mathbf{x})\psi(u|\mathbf{x})\,du+\int_0^1 s(u|\mathbf{x})\psi(u|\mathbf{x})\,du\Big]^2\,dt\\
&\ge c_2\int_0^1 g^2(t|\mathbf{x};\boldsymbol{\tau})\,dt.
\end{aligned}
$$

It follows from Condition 1, (2.1) and (2.2) that there is a positive constant c_3 such that

$$
\sum_i \operatorname{var}\Big(\delta_i g(Y_i|\mathbf{x}_i;\boldsymbol{\tau})-\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\exp g(u|\mathbf{x}_i;\boldsymbol{\theta}^*)\,du\Big)\ge c_3 n I^{-1}|\boldsymbol{\tau}|^2.
$$

This completes the proof of Lemma 3.

Set $\mathbf{I}^*=\mathbf{I}(\boldsymbol{\theta}^*)$. Consider the approximation $\widehat{\varphi}=\widehat{\varphi}_n\in\Theta$ to $\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*$ defined by $\mathbf{I}^*\widehat{\varphi}=\mathbf{S}^*$. Note that $\widehat{\varphi}=(\mathbf{I}^*)^{-1}\mathbf{S}^*$ and hence $E(\widehat{\varphi})=\mathbf{0}$ and $[\mathbf{G}^*(y,\delta,\mathbf{x})]^\top\widehat{\varphi}=[\mathbf{G}^*(y,\delta,\mathbf{x})]^\top(\mathbf{I}^*)^{-1}\mathbf{S}^*$. It follows from (4.10), (2.5) and (4.8) that

(4.14) $\quad |\boldsymbol{\tau}^\top(\mathbf{I}^*)^{-1}\mathbf{G}^*(y,\delta,\mathbf{x})|=O(n^{-1}I|\boldsymbol{\tau}|)\quad$ uniformly in $n,\boldsymbol{\tau}$, y and $\mathbf{x}$.

Lemma 4 $\max_{j\in A}|\widehat{\varphi}_j|=O_p(\sqrt{n^{-1}I\log I})$.

Proof. Since $(\mathbf{I}^*)^{-1}\mathrm{VC}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}$ is the variance-covariance matrix of $\widehat{\varphi}$, it follows from (2.5), (4.8) and Lemma 3 that $\max_j \operatorname{var}(\widehat{\varphi}_j)=O(I/n)$. Observe that

$$\widehat{\varphi}_j = \sum_i [(\mathbf{I}^*)^{-1}\mathbf{G}^*(Y_i, \delta_i, \mathbf{x}_i)]_j.$$

By (4.14),

$$\max_{j\in A} \sup_{y,\mathbf{x}} [(\mathbf{I}^*)^{-1}\mathbf{G}^*(y, \delta, \mathbf{x})]_j = O(I/n).$$

The desired result follows from Condition 5 and Bernstein's inequality.

Proof of (2.12)

The proof is contained in the following lemma.

Lemma 5 (i) $\max_{j\in A} |\widehat{\theta}_j - \theta_j^*|^2 = O_p(n^{-1} I \log I)$.
(ii) $|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* - \widehat{\boldsymbol{\varphi}}|^2 = O_p(n^{-2} I^3 \log I)$.

Proof. It follows from the maximum likelihood equation that [see (4.1)]

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* - \widehat{\boldsymbol{\varphi}} = (\mathbf{I}^*)^{-1}(\mathbf{I}^* + \mathbf{D})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*).$$

According to (2.5) and (4.8),

$$|(\mathbf{I}^*)^{-1}(\mathbf{I}^* + \mathbf{D})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)|^2 = O(n^{-2} I^2 |(\mathbf{I}^* + \mathbf{D})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)|^2).$$

We claim that

$$|(\mathbf{I}^* + \mathbf{D})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)|^2 = O_p\left(n \max_{j\in A} (\widehat{\theta}_j - \theta_j^*)^2\right). \tag{4.15}$$

[The proof of (4.15) will be given shortly.] Therefore

$$|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* - \widehat{\boldsymbol{\varphi}}|^2 = O_p\left(n^{-1} I^2 \max_{j\in A} (\widehat{\theta}_j - \theta_j^*)^2\right).$$

Consequently, by Lemma 4,

$$\max_{j\in A} (\widehat{\theta}_j - \theta_j^*)^2 = O_p\left(n^{-1} I \log I + n^{-1} I^2 \max_{j\in A} (\widehat{\theta}_j - \theta_j^*)^2\right).$$

Thus by Condition 4 ($I = o(n^{1/2-\epsilon})$),

$$\max_{j \in A} (\widehat{\theta}_j - \theta_j^*)^2 = O_p\left(n^{-1} I \log I\right),$$

which yields the desired results.

Proof of (4.15). Set $N_i(t) = \operatorname{ind}(Y_i \le t, \delta_i = 1)$ and $Z_i(t) = \operatorname{ind}(Y_i \ge t)$, $1 \le i \le n$. Thus $N_i(t) = 1$ if and only if the ith subject is observed to have failed by time t and $Z_i(t) = 1$ if and only if the ith subject is still at risk just prior to time t. The log-likelihood function can be written as

$$\ell(\boldsymbol{\theta}) = \sum_i \left(\int g(u|\mathbf{x}_i; \boldsymbol{\theta})\, dN_i(u) - \int Z_i(u) \exp g(u|\mathbf{x}_i; \boldsymbol{\theta})\, du \right).$$

The jth entry of the score function $\mathbf{S}(\boldsymbol{\theta})$ is given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} = \sum_i \left(\int B_j(u|\mathbf{x}_i)\, dN_i(u) - \int B_j(u|\mathbf{x}_i) Z_i(u) \exp g(u|\mathbf{x}_i; \boldsymbol{\theta})\, du \right), \quad j \in A.$$

The entry in row j and column k of the Hessian matrix of $\ell(\boldsymbol{\theta})$ is given by

$$\ell''_{jk}(\boldsymbol{\theta}) = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} = -\sum_i \int_0^1 B_j(u|\mathbf{x}_i) B_k(u|\mathbf{x}_i) Z_i(u) \exp g(u|\mathbf{x}_i; \boldsymbol{\theta})\, du.$$

Set

$$\begin{aligned} \ell'''_{jkm}(\boldsymbol{\theta}) &= \frac{\partial^3 \ell(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_m} \\ &= -\sum_i \int_0^1 B_j(u|\mathbf{x}_i) B_k(u|\mathbf{x}_i) B_m(u|\mathbf{x}_i) Z_i(u) \exp g(u|\mathbf{x}_i; \boldsymbol{\theta})\, du, \end{aligned}$$

where $m = (m_0, m_1, \ldots, m_d)$ with $m_0, m_1, \ldots, m_d \in \{1, \ldots, J\}$. Note that

$$\begin{aligned} &\int_0^1 [\ell''_{jk}(\boldsymbol{\theta}^* + t(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)) - \ell''_{jk}(\boldsymbol{\theta}^*)]\, dt \\ &\qquad = \int_0^1 \left(\int_0^t \sum_m \ell'''_{jkm}(\boldsymbol{\theta}^* + u(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*))(\widehat{\theta}_m - \theta_m^*)\, du \right) dt. \end{aligned}$$

The entry in row j and column k of $\mathbf{I}^* + \mathbf{D}$ can be written as

$$\begin{aligned} &\int_0^1 [\ell''_{jk}(\boldsymbol{\theta}^* + t(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)) - \ell''_{jk}(\boldsymbol{\theta}^*)]\, dt + \ell''_{jk}(\boldsymbol{\theta}^*) - E[\ell''_{jk}(\boldsymbol{\theta}^*)] \\ &\qquad = \sum_m A_{jkm}(\widehat{\theta}_m - \theta_m^*) + \ell''_{jk}(\boldsymbol{\theta}^*) - E[\ell''_{jk}(\boldsymbol{\theta}^*)], \end{aligned}$$

where

$$A_{jkm} = A_{njkm} = \int_0^1 (1-t)\ell'''_{jkm}(\boldsymbol{\theta}^* + t(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*))\,dt.$$

Thus, the jth entry of $(\mathbf{I}^* + \mathbf{D})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is

$$\sum_k \sum_m A_{jkm}(\widehat{\theta}_k - \theta_k^*)(\widehat{\theta}_m - \theta_m^*) + \sum_k \{\ell''_{jk}(\boldsymbol{\theta}^*) - E[\ell''_{jk}(\boldsymbol{\theta}^*)]\}(\widehat{\theta}_k - \theta_k^*).$$

We claim that

$$\sum_j \Big(\sum_k \sum_m A_{jkm}(\widehat{\theta}_k - \theta_k^*)(\widehat{\theta}_m - \theta_m^*)\Big)^2 = O_p\Big(\max_{j\in A} |\widehat{\theta}_j - \theta_j^*|^2 n^2 I^{-2} |\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*|^2\Big) = O_p\Big(n \max_{j\in A} |\widehat{\theta}_j - \theta_j^*|^2\Big) \tag{4.16}$$

and

$$\sum_j \Big(\sum_k \{\ell''_{jk}(\boldsymbol{\theta}^*) - E[\ell''_{jk}(\boldsymbol{\theta}^*)]\}(\widehat{\theta}_k - \theta_k^*)\Big)^2 = O_p\Big(n \max_{j\in A} |\widehat{\theta}_j - \theta_j^*|^2\Big). \tag{4.17}$$

[The proofs of (4.16) and (4.17) will be given shortly.] It follows from (4.16) and (4.17) that

$$|(\mathbf{I}^* + \mathbf{D})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)|^2 = O_p\Big(n \max_{j\in A} (\widehat{\theta}_j - \theta_j^*)^2\Big)$$

as desired.

Proof of (4.16). This follows from (2.9) and the following result.

Lemma 6 *There is a positive constant M_4 such that*

$$\sum_j \Big(\sum_k \sum_m \max_{0\le t\le 1} \Big|\ell'''_{jkm}(\boldsymbol{\theta}^* + t(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\Big|\,|\tau_m|\Big)^2 \le M_4 n^2 I^{-2} |\boldsymbol{\tau}|^2, \quad \boldsymbol{\tau} \in \Theta.$$

Proof. Note that

$$\sum_j \Big(\sum_k \sum_m \max_{0\le t\le 1} \Big| \ell'''_{jkm}(\boldsymbol{\theta}^* + t(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\Big| |\tau_m| \Big)^2$$

$$\le \Big[\max_{0\le t\le 1} \| \exp g(\cdot|\cdot; \boldsymbol{\theta}^* + t(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*))\|_\infty \Big]^2$$

$$\times \sum_j \Big(\sum_i \sum_m |\tau_m| \int B_j(u|\mathbf{x}_i) B_m(u|\mathbf{x}_i)\, du \Big)^2.$$

There is a positive constant J_0 (not depending on n) such that

(4.18) $\quad B_j(u|\mathbf{x}_i) B_m(u|\mathbf{x}_i) = 0 \quad \text{unless } |m_v - j_v| \le J_0, \quad v = 1, \ldots, M.$

Thus, by (2.3) and the basic properties of B-splines,

$$\sum_i \sum_m |\tau_m| \int B_j(u|\mathbf{x}_i) B_m(u|\mathbf{x}_i)\, du$$

$$\le nI^{-1} \sum_{|m_0 - j_0|\le J_0} \sum_{|m_1 - j_1|\le J_0} \cdots \sum_{|m_M - j_M|\le J_0} |\tau_{m_0 m_1 \cdots m_M}|.$$

Hence, by the Schwarz inequality,

$$\sum_j \Big(\sum_i \sum_m |\tau_m| \int B_j(u|\mathbf{x}_i) B_m(u|\mathbf{x}_i)\, du \Big)^2 \le n^2 I^{-2} (2J_0 + 1)^{2d} |\boldsymbol{\tau}|^2.$$

The desired result follows from (2.5), (2.9) and Condition 5.

Proof of (4.17). Set

$$V_{jk}(u) = \sum_i B_j(u|\mathbf{x}_i) B_k(u|\mathbf{x}_i) \{Z_i(u) - E[Z_i(u)]\}\, \lambda^*(u|\mathbf{x}_i).$$

Then $E[V_{jk}(u)] = 0$ for $0 \le u \le 1$ and

$$\ell''_{jk}(\boldsymbol{\theta}^*) - E[\ell''_{jk}(\boldsymbol{\theta}^*)] = -\int V_{jk}(u)\, du.$$

Thus (4.17) follows from the next lemma.

Lemma 7 *Uniformly in $\boldsymbol{\tau} \in \Theta$,*

$$\sum_j \Big(\sum_k |\tau_k| \int V_{jk}(u)\, du \Big)^2 = O_p\Big(n \max_{k_0,\ldots,k_M} \tau^2_{k_0\ldots k_M} \Big).$$

Proof. By (4.18) and the Schwarz inequality,

$$\begin{aligned}
&\sum_j \left(\sum_k |\tau_k| \int V_{jk}(u)\, du \right)^2 \\
&\quad \le \left(\max_{k_0,\ldots,k_M} \tau^2_{k_0\ldots k_M} \right) \sum_j \left(\sum_{|k_0-j_0|\le J_0} \cdots \sum_{|k_M-j_M|\le J_0} \left| \int V_{jk}(u)\, du \right| \right)^2 \\
&\quad \le \left(\max_{k_0,\ldots,k_M} \tau^2_{k_0\ldots k_M} \right) (2J_0+1)^d \\
&\qquad\qquad \times \sum_j \sum_{|k_0-j_0|\le J_0} \cdots \sum_{|k_M-j_M|\le J_0} \int V^2_{jk}(u)\, du.
\end{aligned}$$

Since $E[V^2_{jk}(u)] = \sum_i [B_j(u|\mathbf{x}_i) B_k(u|\mathbf{x}_i) \lambda^*(u|\mathbf{x}_i)]^2 \mathrm{var}(Z_i(u))$ for $0 \le u \le 1$, the desired result follows from (2.3) and (2.5).

This completes the proof of (2.12).

Proof of (2.13)

It follows from Lemma 5(i) and the basic properties of B-splines that $\|\widehat{\phi} - \phi^*\|_\infty = \|\sum_j (\widehat{\theta}_j - \theta^*_j) B_j\|_\infty = O_p(\sqrt{n^{-1} I \log I})$.

Proof of (2.14)

This follows from (4.5), (2.4) and (2.13).

Proof of (2.15)

Set $\Psi(u|\mathbf{x}) = \int_0^1 \exp(\phi^*(u|\mathbf{x}) + w(\widehat{\phi}(u|\mathbf{x}) - \phi^*(u|\mathbf{x})))\, dw$. Note that

$$\begin{aligned}
(4.19) \qquad \widehat{H}(t|\mathbf{x}) - H^*(t|\mathbf{x}) &= \int_0^t (\widehat{\lambda}(u|\mathbf{x}) - \lambda^*(u|\mathbf{x}))\, du \\
&= \int_0^t [\widehat{\phi}(u|\mathbf{x}) - \phi^*(u|\mathbf{x})] \Psi(u|\mathbf{x})\, du
\end{aligned}$$

and

$$(4.20)\quad \widehat{\phi}(t|\mathbf{x}) - \phi^*(t|\mathbf{x}) = \sum_j(\widehat{\theta}_j - \theta_j^* - \widehat{\varphi}_j)B_j(t|\mathbf{x}) + \sum_j \widehat{\varphi}_j B_j(t|\mathbf{x}).$$

It follows from Condition 5, Lemma 5(ii) and the basic properties of B-splines that

$$(4.21)\qquad \int_0^t \Big|\sum_j(\widehat{\theta}_j - \theta_j^* - \widehat{\varphi}_j)B_j(u|\mathbf{x})\Big|\,du = o_p\Big(\sqrt{n^{-1}J^M}\Big).$$

Also, it follows as in the proof of (37) of Stone (1991) that

$$\sup_{0\le t\le 1}\Big|\sum_j \widehat{\varphi}_j \int_0^t B_j(u|\mathbf{x})\lambda^*(u|\mathbf{x})\,du\Big| = O_p\Big(\sqrt{n^{-1}J^M}\Big).$$

The desired result follows from (2.4) and (2.10).

Proof of (2.16)

This follows from (2.6), (2.15) and

$$\begin{aligned}
&\widehat{S}(t|\mathbf{x}) - S^*(t|\mathbf{x})\\
&\quad = \exp(-\widehat{H}(t|\mathbf{x})) - \exp(-H^*(t|\mathbf{x}))\\
&\quad = -[\widehat{H}(t|\mathbf{x}) - H^*(t|\mathbf{x})]\int_0^1 \exp(-H^*(t|\mathbf{x}) - u[\widehat{H}(t|\mathbf{x}) - H^*(t|\mathbf{x})])\,du.
\end{aligned}$$

Proofs of (2.17) and (2.18)

Since $\widehat{f}(t|\mathbf{x}) - f^*(t|\mathbf{x}) = \widehat{\lambda}(t|\mathbf{x})[\widehat{S}(t|\mathbf{x}) - S^*(t|\mathbf{x})] + [\widehat{\lambda}(t|\mathbf{x}) - \lambda^*(t|\mathbf{x})]S^*(t|\mathbf{x})$, the desired results follow from (2.5), (2.7), (2.11), (2.14) and (2.16).

5 Proof of Theorem 3

Throughout this section, we assume that Conditions 1–5 hold.

Lemma 8 $\left|\boldsymbol{\tau}^{\top}\mathrm{VC}(\mathbf{S}^{*})\boldsymbol{\tau}-\boldsymbol{\tau}^{\top}\mathbf{I}^{*}\boldsymbol{\tau}\right| = O\left(nI^{-1}|\boldsymbol{\tau}|^{2}\rho\right), \quad \boldsymbol{\tau}\in\Theta.$

Proof. Recall that the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i}\left(\int g(u|\mathbf{x}_i;\boldsymbol{\theta})\,dN_i(u) - \int Z_i(u)\exp g(u|\mathbf{x}_i;\boldsymbol{\theta})\,du\right)$$

and the score function is given by

$$\begin{aligned}\frac{\partial\ell(\boldsymbol{\theta})}{\partial\theta_j} &= \sum_{i}\left(\int B_j(u|\mathbf{x}_i)\,dN_i(u) - \int B_j(u|\mathbf{x}_i)Z_i(u)\exp g(u|\mathbf{x}_i;\boldsymbol{\theta})\,du\right)\\ &= \sum_{i}\int B_j(u|\mathbf{x}_i)\,dM_i(u|\mathbf{x}_i), \quad j\in A,\end{aligned}$$

where $dM_i(u|\mathbf{x}) = dN_i(u) - Z_i(u)\exp g(u|\mathbf{x};\boldsymbol{\theta})\,du$. Thus,

$$\boldsymbol{\tau}^{\top}\mathrm{VC}(\mathbf{S}^{*})\boldsymbol{\tau} = \mathrm{var}(\boldsymbol{\tau}^{\top}\mathbf{S}^{*}) = \sum_{i}\mathrm{var}\left(\int g(u|\mathbf{x}_i;\boldsymbol{\tau})\,dM_i^{*}(u|\mathbf{x}_i)\right) \tag{5.1}$$

and

$$\begin{aligned}\boldsymbol{\tau}^{\top}\mathbf{I}(\boldsymbol{\theta}^{*})\boldsymbol{\tau} &= \sum_{i}E\left(\int g^{2}(u|\mathbf{x}_i;\boldsymbol{\tau})Z_i(u)\lambda^{*}(u|\mathbf{x}_i)\,du\right)\\ &= \sum_{i}\int g^{2}(u|\mathbf{x}_i;\boldsymbol{\tau})S(u|\mathbf{x}_i)S_C(u|\mathbf{x}_i)\lambda^{*}(u|\mathbf{x}_i)\,du,\end{aligned} \tag{5.2}$$

where $M_i^{*}(t) = N_i(t) - \int_0^t Z_i(u)\exp g(u|\mathbf{x}_i;\boldsymbol{\theta}^{*})\,du$. Let $E^{*}(\cdot)$ and $\mathrm{var}^{*}(\cdot)$ denote the expectation and variance functions taken with respect to f^{*}. According to Theorem 2.5.4 of Fleming and Harrington (1991) (p.77), or Proposition II.4.1 of Andersen et al. (1993) (p.78), $M_i^{*}(u|\mathbf{x}_i)$ is a zero-mean martingale with $\langle M_i^{*}, M_i^{*}\rangle^{*}(t) = \int_0^t Z_i(u)\exp g(u|\mathbf{x}_i;\boldsymbol{\theta}^{*})\,du$. (Here it is necessary to use an alternative probability space with probability measure P^{*}, under which the counting process $N_i(t)$ has an intensity function λ^{*}; $\langle\cdot,\cdot\rangle^{*}$ is the corresponding variation process.) Hence,

$$\begin{aligned}\mathrm{var}^{*}(\boldsymbol{\tau}^{\top}\mathbf{S}^{*}) &= \sum_{i}E^{*}\left(\int g^{2}(u|\mathbf{x}_i;\boldsymbol{\tau})Z_i(u)\lambda^{*}(u|\mathbf{x}_i)\,du\right)\\ &= \sum_{i}\int g^{2}(u|\mathbf{x}_i;\boldsymbol{\tau})S^{*}(u|\mathbf{x}_i)S_C(u|\mathbf{x}_i)\lambda^{*}(u|\mathbf{x}_i)\,du.\end{aligned} \tag{5.3}$$

It follows from (2.5), (5.2), (5.3) (2.1) and (2.2) that

$$(5.4)\qquad \left|\boldsymbol{\tau}^\top \mathbf{I}(\boldsymbol{\theta}^*)\boldsymbol{\tau} - \mathrm{var}^*(\boldsymbol{\tau}^\top \mathbf{S}^*)\right| = O\Big(nI^{-1}|\boldsymbol{\tau}|^2\|S^* - S\|_\infty\Big), \quad \boldsymbol{\tau} \in \Theta.$$

Set

$$\begin{aligned} U_i &= \int g(u|\mathbf{x}_i;\boldsymbol{\tau})\, dM_i^*(u|\mathbf{x}_i) \\ &= \delta_i g(Y_i|\mathbf{x}_i;\boldsymbol{\tau}) - \int g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i) Z_i(u)\, du. \end{aligned}$$

Then

$$(5.5)\quad \mathrm{var}(U_i) - \mathrm{var}^*(U_i) = E(U_i^2) - [E(U_i)]^2 - E^*(U_i^2) + [E^*(U_i)]^2.$$

Write $g = g(\cdot;\boldsymbol{\tau})$. Then

$$\begin{aligned} E(U_i) &= E\Big(\delta_i g(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\Big) - E\left(\int g\lambda^* Z_i\right) \\ &= \int S_C(u|\mathbf{x}_i) g(u|\mathbf{x}_i) f(u|\mathbf{x}_i) - \int g(u|\mathbf{x}_i)\lambda^*(u|\mathbf{x}_i) S_C(u|\mathbf{x}_i) S(u|\mathbf{x}_i) \end{aligned}$$

and

$$\begin{aligned} E^*(U_i) &= E^*\Big(\delta_i g(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\Big) - E^*\left(\int g\lambda^* Z_i\right) \\ &= \int S_C(u|\mathbf{x}_i) g(u|\mathbf{x}_i) f^*(u|\mathbf{x}_i) - \int g(u|\mathbf{x}_i)\lambda^*(u|\mathbf{x}_i) S_C(u|\mathbf{x}_i) S^*(u|\mathbf{x}_i). \end{aligned}$$

Thus,

$$\begin{aligned} [E(U_i)]^2 - [E(U_i)]^2 &= [E(U_i) - E^*(U_i)][E(U_i) + E^*(U_i)] \\ &= \left(\int S_C g(f - f^*) - \int g\lambda^* S_C(S - S^*)\right) \\ &\qquad \times \left(\int S_C g(f^* + f) - \int g\lambda^* S_C(S^* + S)\right). \end{aligned}$$

Hence, by (2.2), (2.5), (2.7), (2.8) and (2.1),

$$(5.6)\quad \sum_i \left|[E(U_i)]^2 - [E(U_i)]^2\right| = O\Big(nI^{-1}|\boldsymbol{\tau}|^2\|f^* - f\|_\infty\Big), \quad \boldsymbol{\tau} \in \Theta.$$

We claim that

$$(5.7)\qquad \sum_i \left|E(U_i^2) - E^*(U_i^2)\right| = O\Big(nI^{-1}|\boldsymbol{\tau}|^2\|f^* - f\|_\infty\Big), \quad \boldsymbol{\tau} \in \Theta.$$

[The proof of (5.7) will be given shortly.] Hence, by (5.1) and (5.5)–(5.7),

$$(5.8)\qquad \left|\boldsymbol{\tau}^\top \mathrm{VC}(\mathbf{S}^*)\boldsymbol{\tau} - \mathrm{var}^*(\boldsymbol{\tau}^\top \mathbf{S}^*)\right| = O\Big(nI^{-1}|\boldsymbol{\tau}|^2\|f^* - f\|_\infty\Big), \quad \boldsymbol{\tau} \in \Theta.$$

The desired result follows from (2.7), (2.8), (5.4) and (5.8).

To verify (5.7), we first note that,

$$(5.9)\qquad \begin{aligned} E(U_i^2) = {} & E\Big(\delta g^2(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\Big) + E\left[\left(\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i)\,du\right)^2\right] \\ & - 2E\left(\delta g(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i)\,du\right) \end{aligned}$$

and

$$(5.10)\qquad \begin{aligned} E^*(U_i^2) = {} & E^*\Big(\delta g^2(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\Big) + E^*\left[\left(\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i)\,du\right)^2\right] \\ & - 2E^*\left(\delta g(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i)\,du\right). \end{aligned}$$

By (2.1) and (2.2),

$$(5.11)\qquad \begin{aligned} & \sum_i \left|E\Big(\delta g^2(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\Big) - E^*\Big(\delta g^2(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\Big)\right| \\ & \quad \le \sum_i \int S_C(u|\mathbf{x}_i)g^2(u|\mathbf{x}_i)\left|f(u|\mathbf{x}_i) - f^*(u|\mathbf{x}_i)\right| du \\ & \quad = O\Big(nI^{-1}|\boldsymbol{\tau}|^2\|f - f^*\|_\infty\Big). \end{aligned}$$

Set $S_Y(y|\mathbf{x}) = S(y|\mathbf{x})S_C(y|\mathbf{x})$ and $S_Y^*(y|\mathbf{x}) = S^*(y|\mathbf{x})S_C(y|\mathbf{x})$. Then

$$dS_Y(y|\mathbf{x}) = S(y|\mathbf{x})dS_C(y|\mathbf{x}) - f(y|\mathbf{x})S_C(y|\mathbf{x}),$$

$$dS_Y^*(y|\mathbf{x}) = S^*(y|\mathbf{x})dS_C(y|\mathbf{x}) - f^*(y|\mathbf{x})S_C(y|\mathbf{x})$$

and

$$E\left[\left(\int_0^Y g(u|\mathbf{x};\boldsymbol{\tau})\lambda^*(u|\mathbf{x})\,du\right)^2\right] - E^*\left[\left(\int_0^Y g(u|\mathbf{x};\boldsymbol{\tau})\lambda^*(u|\mathbf{x})\,du\right)^2\right]$$
$$= \int \left(\int_0^y g(u|\mathbf{x};\boldsymbol{\tau})\lambda^*(u|\mathbf{x})\,du\right)^2 dS_Y^*(y|\mathbf{x})$$
$$- \int \left(\int_0^y g(u|\mathbf{x})\lambda^*(u|\mathbf{x})\,du\right)^2 dS_Y(y|\mathbf{x})$$
$$= \int \left(\int_0^y g(u|\mathbf{x})\lambda^*(u|\mathbf{x})\,du\right)^2 \Big(S_C(y|\mathbf{x})[f(y|\mathbf{x}) - f^*(y|\mathbf{x})]\,dy$$
$$- [S(y|\mathbf{x}) - S^*(y|\mathbf{x})]\,dS_C(y|\mathbf{x})\Big).$$

Thus, by (2.5), (2.7), (2.1) and (2.2),

$$(5.12)\quad \sum_i \left|E\left[\left(\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i)\,du\right)^2\right]\right.$$
$$\left. - E^*\left[\left(\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i)\,du\right)^2\right]\right| = O\Big(nI^{-1}|\boldsymbol{\tau}|^2\|f - f^*\|_\infty\Big).$$

Also,

$$E\left(\delta g(Y|\mathbf{x};\boldsymbol{\tau})\int_0^Y g(u|\mathbf{x};\boldsymbol{\tau})\lambda^*(u|\mathbf{x})\right)$$
$$- E^*\left(\delta g(Y|\mathbf{x};\boldsymbol{\tau})\int_0^Y g(u|\mathbf{x};\boldsymbol{\tau})\lambda^*(u|\mathbf{x})\,du\right)$$
$$= \int S_C(t|\mathbf{x})g(t|\mathbf{x};\boldsymbol{\tau})[f(t|\mathbf{x}) - f^*(t|\mathbf{x})]$$
$$\times \left(\int_0^t g(u|\mathbf{x};\boldsymbol{\tau})\lambda^*(u|\mathbf{x})\,du\right) dt.$$

Thus, by the Schwarz inequality, (2.5), (2.1) and (2.2),

$$(5.13)\quad \sum_i \left|E\left(\delta_i g(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i)\,du\right)\right.$$
$$\left. -E^*\left(\delta g(Y_i|\mathbf{x}_i;\boldsymbol{\tau})\int_0^{Y_i} g(u|\mathbf{x}_i;\boldsymbol{\tau})\lambda^*(u|\mathbf{x}_i)\,du\right)\right|$$
$$= O\Big(nI^{-1}|\boldsymbol{\tau}|^2\|f - f^*\|_\infty\Big).$$

It follows from (5.9)–(5.13) that (5.7) holds.

This completes the proof of Lemma 8.

Proof of (2.19)

Set $W_i = [\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1}\mathbf{G}^*(Y_i|\mathbf{x}_i)$, $i = 1, \ldots, n$. Observe that

$$[\mathbf{B}(t|\mathbf{x})]^\top \widehat{\varphi} = \sum_j \widehat{\varphi}_j B_j(t|\mathbf{x}) = \sum_i W_i.$$

Also, $E([\mathbf{B}(t|\mathbf{x})]^\top \widehat{\varphi}) = 0$ and

$$\operatorname{var}([\mathbf{B}(t|\mathbf{x})]^\top \widehat{\varphi}) = [\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1} \mathrm{VC}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\mathbf{B}(t|\mathbf{x}).$$

Lemma 9 $\operatorname{var}([\mathbf{B}(t|\mathbf{x})]^\top \widehat{\varphi}) \sim n^{-1}I$.

Proof. Since $B_j \geq 0$ and $\sum_j B_j = 1$, we have that

(5.14) $$|\mathbf{B}(t|\mathbf{x})| \sim 1.$$

By (5.14), (4.8) and (2.5),

(5.15) $$|(\mathbf{I}^*)^{-1}\mathbf{B}(t|\mathbf{x})| = O\Big(n^{-1}I\Big).$$

By Lemma 8, (5.15) and Condition 5 [$\rho = o(1)$],

(5.16) $$\Big|[\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1} \mathrm{VC}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\mathbf{B}(t|\mathbf{x}) - [\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1}\mathbf{B}(t|\mathbf{x})\Big| = o\Big(n^{-1}I\Big).$$

By (5.14) and (4.7),

(5.17) $$[\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1}\mathbf{B}(t|\mathbf{x}) \sim n^{-1}I.$$

It follows from (5.16) and (5.17) that

$$\operatorname{var}([\mathbf{B}(t|\mathbf{x})]^\top \widehat{\varphi}) = (1 + o(1))[\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1}\mathbf{B}(t|\mathbf{x}) \sim n^{-1}I$$

as desired.

Lemma 10

$$\frac{\sum_j \widehat{\varphi}_j B_j(t|\mathbf{x})}{\mathrm{SD}([\mathbf{B}(t|\mathbf{x})]^\top \widehat{\varphi})} \xrightarrow{d} N(0,1).$$

Proof. The random variables $W_1, \ldots, W_n$ are independent with mean zero. Moreover, by (5.15) and (4.10),

$$|W_i|^2 = |[\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1} \mathbf{G}^*(Y_i|\mathbf{x}_i)|^2 = O(I^2/n^2). \tag{5.18}$$

The desired result follows from Condition 5, Lemma 9 and the central limit theorem.

Now according to Lemma 5(ii) and Condition 5,

$$|[\mathbf{B}(t|\mathbf{x})]^\top (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* - \widehat{\boldsymbol{\varphi}})| = o_p\left(\sqrt{n^{-1}I}\right).$$

Since $\widehat{\phi}(t|\mathbf{x}) - \phi^*(t|\mathbf{x}) = \sum_j (\widehat{\theta}_j - \theta_j^*) B_j(t|\mathbf{x})$, we now conclude from Lemmas 9 and 10 that

$$\frac{\widehat{\phi}(t|\mathbf{x}) - \phi^*(t|\mathbf{x})}{\mathrm{SD}([\mathbf{B}(t|\mathbf{x})]^\top \widehat{\boldsymbol{\varphi}})} \xrightarrow{d} N(0,1).$$

By (5.17),

$$\mathrm{AV}(\widehat{\phi}(t|\mathbf{x})) = [\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1} \mathbf{B}(t|\mathbf{x}) \sim n^{-1} I.$$

Thus, by (5.16),

$$\mathrm{var}([\mathbf{B}(t|\mathbf{x})]^\top \widehat{\boldsymbol{\varphi}}) \simeq \mathrm{AV}(\widehat{\phi}(t|\mathbf{x})).$$

Hence,

$$\frac{\widehat{\phi}(t|\mathbf{x}) - \phi^*(t|\mathbf{x})}{\mathrm{ASD}(\widehat{\phi}(t|\mathbf{x}))} \xrightarrow{d} N(0,1).$$

This completes the proof of the first part of (2.19).

To prove the second part of (2.19), set $\widehat{\mathbf{I}} = \mathbf{I}(\widehat{\boldsymbol{\theta}})$. The proof depends on the next two lemmas.

Lemma 11 *Uniformly in $\boldsymbol{\tau} \in \Theta$,*

$$\left|(\widehat{\mathbf{I}} - \mathbf{I}^*)\boldsymbol{\tau}\right|^2 = O_p\left(n|\boldsymbol{\tau}|^2 I^{-1} \log I\right).$$

Proof. Observe that

$$\begin{aligned} &E[\ell''_{jk}(\boldsymbol{\theta})] - E\ell''_{jk}(\boldsymbol{\theta}^*)] \\ &\quad \le \max_{m\in A} |\theta_m - \theta_m^*| \left[\max_{0\le t\le 1} \|\exp g(\cdot|\cdot; \boldsymbol{\theta}^* + t(\boldsymbol{\theta} - \boldsymbol{\theta}^*))\|_\infty\right] \\ &\quad \times \sum_m \sum_i \int B_j(u|\mathbf{x}_i) B_k(u|\mathbf{x}_i) B_m(u|\mathbf{x}_i) S(u|\mathbf{x}_i) S_C(u|\mathbf{x}_i)\, du, \end{aligned}$$

It follows from the basic properties of B-splines as in the proof of Lemma 6 that uniformly in $\boldsymbol{\theta}, \boldsymbol{\tau} \in \Theta$,

$$\sum_j \Big[\sum_k \Big(E[\ell''_{jk}(\boldsymbol{\theta})] - E\ell''_{jk}(\boldsymbol{\theta}^*)]\Big)\tau_k\Big]^2$$
$$= O\bigg(n^2 \max_{m\in A}(\theta_m - \theta^*_m)^2\Big[\max_{0\le t\le 1} \|\exp g(\cdot|\cdot;\boldsymbol{\theta}^* + t(\boldsymbol{\theta}-\boldsymbol{\theta}^*))\|_\infty\Big]^2 I^{-2}|\boldsymbol{\tau}|^2\bigg).$$

The desired conclusion follows from (2.5), (2.9), Condition 5 and (2.12). This completes the proof of Lemma 11.

Lemma 12 *Uniformly in* $\boldsymbol{\tau} \in \Theta$,

$$\Big|(\widehat{\mathbf{I}}^{-1} - (\mathbf{I}^*)^{-1})\boldsymbol{\tau}\Big|^2 = O_p\Big(n^{-3}I^3(\log I)|\boldsymbol{\tau}|^2\Big).$$

Proof. Since $\widehat{\mathbf{I}}^{-1} - (\mathbf{I}^*)^{-1} = (\mathbf{I}^*)^{-1}(\mathbf{I}^* - \widehat{\mathbf{I}})\widehat{\mathbf{I}}^{-1}$, the desired result follows from (2.5), (2.14), (4.8) [with $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ and $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$] and Lemma 11.

The proof of the second part of (2.19) will now be given. Recall that

$$\mathrm{SE}(\widehat{\phi}(t|\mathbf{x})) = \sqrt{[\mathbf{B}(t|\mathbf{x})]^\top \widehat{\mathbf{I}}^{-1}\mathbf{B}(t|\mathbf{x})}$$

and

$$\mathrm{ASD}(\widehat{\phi}(t|\mathbf{x})) = \sqrt{[\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1}\mathbf{B}(t|\mathbf{x})}.$$

By the Schwarz inequality, (5.14), Lemma 12 and Condition 5,

$$\begin{aligned}
&\Big|[\mathbf{B}(t|\mathbf{x})]^\top \widehat{\mathbf{I}}^{-1}\mathbf{B}(t|\mathbf{x}) - [\mathbf{B}(t|\mathbf{x})]^\top (\mathbf{I}^*)^{-1}\mathbf{B}(t|\mathbf{x})\Big| \\
&\quad = |[\mathbf{B}(t|\mathbf{x})]^\top [\widehat{\mathbf{I}}^{-1} - (\mathbf{I}^*)^{-1}]\mathbf{B}(t|\mathbf{x})| \\
&\quad \le |\mathbf{B}(t|\mathbf{x})||[\widehat{\mathbf{I}}^{-1} - (\mathbf{I}^*)^{-1}]\mathbf{B}(t|\mathbf{x})| \\
&\quad = O_p\Big(\sqrt{n^{-3}I^3\log I}\Big). \\
&\quad = o_p\Big(n^{-1}I\Big).
\end{aligned}$$

It now follows from (5.17) that the second part of (2.19) holds.

Proof of (2.20)

It follows from (2.4) and (4.5) that

$$\widehat{\lambda}(t|\mathbf{x}) - \lambda^*(t|\mathbf{x}) = [\widehat{\phi}(t|\mathbf{x}) - \phi^*(t|\mathbf{x})]\lambda^*(t|\mathbf{x}) + o_p\Big(\sqrt{n^{-1}I}\Big).$$

By (2.5) and (5.17), $\mathrm{ASD}(\widehat{\lambda}(t|\mathbf{x})) = \lambda^*(t|\mathbf{x})\mathrm{ASD}(\widehat{\phi}(t|\mathbf{x})) \sim \sqrt{n^{-1}I}$. Thus the desired result follows from $\mathrm{SE}(\widehat{\lambda}(t|\mathbf{x})) = \widehat{\lambda}(t|\mathbf{x})\,\mathrm{SE}(\widehat{\phi}(t|\mathbf{x}))$, (5.16), (2.14) and (2.19).

Proof of (2.21)

Set $\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x}) = \int_0^t \lambda^*(u|\mathbf{x})\,du$, $0 \le t \le 1$. Then

(5.19) $$\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x}) = \int_0^t \lambda^*(u|\mathbf{x})\mathbf{B}(u|\mathbf{x})\,du.$$

It follows from (2.4) and (4.19)–(4.21) that

(5.20) $$\widehat{H}(t|\mathbf{x}) - H^*(t|\mathbf{x}) = [\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x})]^{\top}\widehat{\varphi} + o_p\Big(\sqrt{n^{-1}J^M}\Big).$$

Now observe that

(5.21) $$\begin{aligned}&\mathrm{var}\Big([\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x})]^{\top}\widehat{\varphi}\Big)\\ &\quad= [\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x})]^{\top}(\mathbf{I}^*)^{-1}\mathrm{VC}(\mathbf{S}^*)(\mathbf{I}^*)^{-1}\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x}).\end{aligned}$$

By (5.19), (4.9), (2.5) and the basic properties of B-splines,

(5.22) $$|\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x})|^2 \sim J^{-1}, \quad t > 0.$$

It follows from (5.21), (5.22), (2.5), (4.8) and Lemma 8 that

(5.23) $$\begin{aligned}&\mathrm{var}\Big([\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x})]^{\top}\widehat{\varphi}\Big)\\ &\quad= [\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x})]^{\top}(\mathbf{I}^*)^{-1}\nabla\Gamma(\boldsymbol{\theta}^*; t, \mathbf{x}) + O\Big(n^{-1}J^M\rho\Big).\end{aligned}$$

The following result is an easy consequence of (4.7) and (5.22).

Lemma 13 $[\nabla\Gamma(\boldsymbol{\theta}^*;t,\mathbf{x})]^\top(\mathbf{I}^*)^{-1}\nabla\Gamma(\boldsymbol{\theta}^*;t,\mathbf{x}) \sim n^{-1}J^M, \quad t>0.$

The next result establishes the central limit theorem for $[\nabla\Gamma(\boldsymbol{\theta}^*;t,\mathbf{x})]^\top\widehat{\varphi}$. The proof is similar to the argument for Lemma 10.

Lemma 14 *If* $t>0$, *then*

$$\frac{[\nabla\Gamma(\boldsymbol{\theta}^*;t,\mathbf{x})]^\top\widehat{\varphi}}{\mathrm{SD}([\nabla\Gamma(\boldsymbol{\theta}^*;t,\mathbf{x})]^\top\widehat{\varphi})} \xrightarrow{d} N(0,1).$$

Since $\mathrm{ASD}(\widehat{H}(t|\mathbf{x})) = \sqrt{[\nabla\Gamma(\boldsymbol{\theta}^*;t,\mathbf{x})]^\top(\mathbf{I}^*)^{-1}\nabla\Gamma(\boldsymbol{\theta}^*;t,\mathbf{x})}$, we conclude from (5.20), (5.23), and Lemmas 13 and 14 that the first part of (2.21) is valid. By (5.19), (2.14), and (4.9) with $\boldsymbol{\theta}^*$ replaced by $\widehat{\boldsymbol{\theta}}$,

$$|\nabla\Gamma(\widehat{\boldsymbol{\theta}};t,\mathbf{x})|^2 = O_p\big(J^{-1}\big). \tag{5.24}$$

Similarly,

$$\Big|\nabla\Gamma(\widehat{\boldsymbol{\theta}};t,\mathbf{x}) - \nabla\Gamma(\boldsymbol{\theta}^*;t,\mathbf{x})\Big|^2 = o_p\big(J^{-1}\big). \tag{5.25}$$

Since $\mathrm{SE}(\widehat{H}(t|\mathbf{x})) = \sqrt{\nabla\Gamma(\widehat{\boldsymbol{\theta}};t,\mathbf{x})]^\top\widehat{\mathbf{I}}^{-1}\nabla\Gamma(\widehat{\boldsymbol{\theta}};t,\mathbf{x})}$, we conclude from (4.8), (5.22), (5.24), (5.25), Lemmas 12 and 13, Condition 5 and the Schwarz inequality that the second part of (2.21) is valid.

Proof of (2.22)

By a similar expansion as in the proof of (2.16),

$$\widehat{S}(t|\mathbf{x}) - S^*(t|\mathbf{x}) = -[\widehat{H}(t|\mathbf{x}) - H^*(t|\mathbf{x})]S^*(t|\mathbf{x}) + o_p\Big(\sqrt{n^{-1}J^M}\Big).$$

According to Lemma 13 and (2.7), $\mathrm{ASD}(\widehat{S}(t|\mathbf{x})) = S^*(t|\mathbf{x})\,\mathrm{ASD}(\widehat{H}(t|\mathbf{x})) \sim \sqrt{n^{-1}J^M}$. We conclude from $\mathrm{SE}(\widehat{S}(t|\mathbf{x})) = \widehat{S}(t|\mathbf{x})\,\mathrm{SE}(\widehat{H}(t|\mathbf{x}))$, (2.16) and (2.21) that (2.22) holds.

Proof of (2.23)

By (2.5), (2.14) and (2.16),

$$\text{(5.26)} \qquad \begin{aligned} &\widehat{f}(t|\mathbf{x}) - f^*(t|\mathbf{x}) \\ &= (\widehat{\lambda}(t|\mathbf{x}) - \lambda^*(t|\mathbf{x}))S^*(t|\mathbf{x}) + (\widehat{S}(t|\mathbf{x}) - S^*(t|\mathbf{x}))\widehat{\lambda}(t|\mathbf{x}) \\ &= [\widehat{\lambda}(t|\mathbf{x}) - \lambda^*(t|\mathbf{x})]S^*(t|\mathbf{x}) + O_p\Big(\sqrt{n^{-1}J^M}\Big). \end{aligned}$$

By (5.22) and (5.24), respectively,

$$\text{(5.27)} \qquad \Big|\int_0^t \mathbf{B}\lambda^*\Big|^2 = O\big(J^{-1}\big)$$

and

$$\text{(5.28)} \qquad \Big|\int_0^t \mathbf{B}\widehat{\lambda}\Big|^2 = O_p\big(J^{-1}\big).$$

It follows from (4.7), (5.14), (5.27), (5.28), Lemma 11, Condition 5 and the Schwarz inequality that

$$\begin{aligned} \mathrm{ASD}(\widehat{f}(t|\mathbf{x})) &= f^*(t|\mathbf{x})\sqrt{\Big(\mathbf{B}(t|\mathbf{x}) - \int_0^t \mathbf{B}\lambda^*\Big)^{\top}(\mathbf{I}^*)^{-1}\Big(\mathbf{B}(t|\mathbf{x}) - \int_0^t \mathbf{B}\lambda^*\Big)} \\ &\simeq f^*(t|\mathbf{x})\,\mathrm{ASD}(\widehat{\phi}(t|\mathbf{x})) \end{aligned}$$

and

$$\begin{aligned} \mathrm{SE}(\widehat{f}(t|\mathbf{x})) &= \widehat{f}(t|\mathbf{x})\sqrt{\Big(\mathbf{B}(t|\mathbf{x}) - \int_0^t \mathbf{B}(t|\mathbf{x})\widehat{\lambda}\Big)^{\top}\widehat{\mathbf{I}}^{-1}\Big(\mathbf{B}(t|\mathbf{x}) - \int_0^t \mathbf{B}\widehat{\lambda}\Big)} \\ &\simeq \widehat{f}(t|\mathbf{x})\,\mathrm{SE}(\widehat{\phi}(t|\mathbf{x})). \end{aligned}$$

The desired result follows from (5.26), (2.18), (2.19), (2.20) and $\mathrm{ASD}(\widehat{f}(t|\mathbf{x})) \sim \sqrt{I/n}$.

6 Appendix. Proof of (2.4)

Let $M_1, M_2, \ldots$ denote constants greater than 1. According to Conditions 1 and 2,

$$\text{(6.1)} \qquad M_1^{-1} \le S(t|\mathbf{x})S_C(t|\mathbf{x}) \le M_1, \qquad (t, \mathbf{x}) \in \mathcal{T} \times \mathcal{X},$$

and $S(t|\mathbf{x})S_C(t|\mathbf{x}) = 0$ for $t > 1$.

Let $\mathcal{A}$ denote a collection of functions ϕ on $\mathcal{T} \times \mathcal{X}$ satisfying the Hölder condition

$$|\phi(\mathbf{z}) - \phi(\mathbf{z}_0)| \leq \gamma|\mathbf{z} - \mathbf{z}_0|^\beta, \qquad \mathbf{z}, \mathbf{z}_0 \in \mathcal{T} \times \mathcal{X},$$

and the boundedness condition

$$(6.2) \qquad \|\phi\|_\infty \leq M_2, \qquad \phi \in \mathcal{A}.$$

Note that if $\phi \in \mathcal{A}$ and $0 \leq u < 1$, then $u\phi \in \mathcal{A}$. Set

$$\rho = \rho(\phi) = \inf_{g \in G} \|\phi - g\|_\infty, \qquad \phi \in \mathcal{A},$$

and note that $\rho(\phi) \leq M_2$ for $\phi \in \mathcal{A}$. Writing ϕ^* as $Q\phi$ and closely following the argument in Stone (1989), we will obtain an inequality of the form

$$(6.3) \qquad \|\phi - Q\phi\|_\infty \leq M\rho(\phi), \qquad \phi \in \mathcal{A},$$

where the positive constant M depends on $\mathcal{A}$ and the degree m of G, but not on the dimension I of G. We conclude from (6.3) that (2.4) holds.

Let ψ be a function on $\mathcal{T} \times \mathcal{X}$ such that

$$(6.4) \qquad M_3^{-1} \leq \psi(t|\mathbf{x}) \leq M_3, \qquad (t, \mathbf{x}) \in \mathcal{T} \times \mathcal{X}.$$

Consider the $I \times I$ matrix $\mathbf{M}$ whose (j,l)th entry is $\sum_i \int B_j(t|\mathbf{x}_i)B_l(t|\mathbf{x}_i)\psi(t|\mathbf{x}_i)\, dt$ for j and l range over A. It follows from (2.1) and (2.2) that $\mathbf{M}$ is invertible. Let γ_{jl} denote the (j,l)th entry of $\mathbf{M}^{-1}$. Then $\|\mathbf{M}^{-1}\|_\infty = \max_j \sum_l |\gamma_{jl}|$. By a slight extension of a result in de Boor (1976), $\|\mathbf{M}^{-1}\|_\infty \leq M_4 n^{-1} I$. This has the following consequence.

Lemma 15 *Set* $g = \sum_j \theta_j B_j$. *Then*

$$\max_j |\theta_j| \leq M_4 n^{-1} I \max_j \left| \sum_i \int g(t|\mathbf{x}_i) B_j(t|\mathbf{x}_i)\psi(t|\mathbf{x}_i)\, dt \right|.$$

We now verify (6.3). Choose $\phi \in \mathcal{A}$ and $g \in G$. Since

$$\sum_i \int \Big\{[ug(t|\mathbf{x}_i) + Q\phi(t|\mathbf{x}_i)] \exp\phi(t|\mathbf{x}_i) \\ - \exp(ug(t|\mathbf{x}_i) + Q\phi(t|\mathbf{x}_i))\Big\} S(t|\mathbf{x}_i) S_C(t|\mathbf{x}_i)\, dt$$

is maximized at $u = 0$,

$$\sum_i \int g(t|\mathbf{x}_i)[\exp\phi(t|\mathbf{x}_i) - \exp Q\phi(t|\mathbf{x}_i)] S(t|\mathbf{x}_i) S_C(t|\mathbf{x}_i)\, dt = 0.$$

Consequently, for $j \in A$,

$$\sum_i \int B_j(t|\mathbf{x}_i)[\exp\phi(t|\mathbf{x}_i) - \exp Q\phi(t|\mathbf{x}_i)] S(t|\mathbf{x}) S_C(t|\mathbf{x}_i)\, dt = 0. \tag{6.5}$$

Let $\phi \in \mathcal{A}$. Then there is an $\bar{\phi} \in G$ such that $\|\phi - \bar{\phi}\|_\infty = \rho(\phi)$. Note that $Q\bar{\phi} = \bar{\phi}$. Note also that $\|\bar{\phi}\|_\infty \le 2M_2$ and hence that $\exp(-2M_2) \le \exp\bar{\phi} \le \exp(2M_2)$ and

$$\|\exp\bar{\phi} - \exp\phi\|_\infty \le \exp(2M_2)\rho(\phi). \tag{6.6}$$

By (6.1), (2.1), (2.3), (6.5) and (6.6),

$$\Big|\sum_i \int B_j(t|\mathbf{x}_i)[\exp\bar{\phi}(t|\mathbf{x}_i) - \exp Q\phi(t|\mathbf{x}_i)] S(t|\mathbf{x}_i) S_C(t|\mathbf{x}_i)\, dt\Big| \\ \le M_1 M_5 n I^{-1} \exp(2M_2)\rho(\phi), \quad j \in A. \tag{6.7}$$

Write $Q\phi - \bar{\phi} = \sum_j \theta_j B_j$ and set $\epsilon = \max_j |\theta_j|$. Now $\|Q\phi - \bar{\phi}\|_\infty \le \epsilon$ and hence

$$\|\phi - Q\phi\|_\infty \le \epsilon + \rho(\phi). \tag{6.8}$$

By repeatedly applying (viii) on Page 155 of de Boor (1978), there is a positive constant M_6, depending only on q, such that

$$\epsilon \le M_6 \|Q\phi - \bar{\phi}\|_\infty. \tag{6.9}$$

Since $\exp Q\phi = (\exp\bar{\phi})\exp(\sum_j \theta_j B_j)$,

$$\Big\|\exp Q\phi - \exp\bar{\phi} - (\exp\bar{\phi})\sum_j \theta_j B_j\Big\|_\infty \le \exp(2M_2 + \epsilon)\frac{\epsilon^2}{2}.$$

We now conclude from from (2.1), (2.2) and (6.7) that, for $j \in A$,

$$\left|\sum_i \int B_j(t|\mathbf{x}_i) \sum_l \theta_l B_l(t|\mathbf{x}_i) \exp \bar{\phi}(t|\mathbf{x}_i) S(t|\mathbf{x}_i) S_C(t|\mathbf{x}_i)\, dt\right|$$
$$\leq \; M_1 M_5 n I^{-1} \exp(2M_2)\Big(\rho(\phi) + \exp(\epsilon)\frac{\epsilon^2}{2}\Big).$$

According to (6.4), (6.7) and Lemma 16 applied to $\psi = SS_C \exp \bar{\phi}$ [with $M_3 = M_1 \exp(2M_2)$],

$$\epsilon \leq M_1 M_4 M_5 \exp(2M_2)\Big(\rho(\phi) + \exp(\epsilon)\frac{\epsilon^2}{2}\Big).$$

Suppose now that

$$M_1 M_4 M_5 \exp(2M_2 + \epsilon)\epsilon \leq 1. \tag{6.10}$$

Then $\epsilon \leq 2M_1 M_4 M_5 \exp(2M_2)\rho(\phi)$ and hence, by (6.8),

$$\|\phi - Q\phi\|_\infty \leq M_7 \rho(\phi), \tag{6.11}$$

where $M_7 = 2[\, M_1 M_4 M_5 \exp(2M_2) + 1\,]$. According to (6.9), a sufficient condition for (6.10) and hence for (6.11) is

$$\|Q\phi - \bar{\phi}\|_\infty \leq M_8^{-1}. \tag{6.12}$$

Let $0 < \rho_0 < 2^{-1} M_7^{-1} M_8^{-1}$. There is a positive integer I_0, depending on M_1 and the degree of G, such that

$$\rho(\phi) \leq \rho_0, \qquad I \geq I_0 \text{ and } \phi \in \mathcal{A} \tag{6.13}$$

[see Theorem 12.8 of Schumaker (1981)]. Let $I \geq I_0$. Suppose that

$$\|\phi - Q\phi\|_\infty \leq 2^{-1} M_8^{-1}. \tag{6.14}$$

Since $\|\phi - \bar{\phi}\|_\infty = \rho(\phi) \leq 2^{-1} M_8^{-1}$, (6.12) holds.

We will now verify that (6.14) necessarily holds for $I \geq I_0$. Suppose not. Now $\|u\phi - Q(u\phi)\|_\infty$ is continuous in u for $0 \leq u < 1$ (since the expected log-likelihood is a strictly concave function of $\theta_1, \ldots, \theta_I$ and it is continuous in $u, \theta_1, \ldots, \theta_I$) and it approaches 0 as $u \to 0$. Thus there is a value of $u \in (0,1)$ such that $\|u\phi - Q(u\phi)\|_\infty = 2^{-1}M_8^{-1}$. By the previous argument, (6.11) and (6.13) hold with ϕ replaced by $u\phi$; hence

$$\|u\phi - Q(u\phi)\|_\infty \le M_7\rho(u\phi) \le M_7\rho_0 < 2^{-1}M_8^{-1},$$

which yields a contradiction.

We have now shown that

$$\|\phi - Q\phi\|_\infty \le M_8\rho(\phi), \qquad I \ge I_0 \text{ and } \phi \in \mathcal{A}.$$

To complete the proof of (6.3) we need to show that

$$\|\bar{\phi} - Q\phi\|_\infty \le M_9\rho(\phi), \qquad I < I_0 \text{ and } \phi \in \mathcal{A}.$$

But this result, for each I, follows in a straightforward manner by a compactness argument.

REFERENCES

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. AND KEIDING, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

DE BOOR, C. (1976). A bound on the L_∞-norm of the L_2 approximation by splines in terms of a global mesh ratio. *Math. Comp.* **30** 365–771.

DE BOOR, C. (1978). *A Practical Guide to Splines.* Springer-Verlag, New York.

BRESLOW, N. E. AND CROWLEY, J. J. (1974). A large sample study of life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation.* Springer-Verlag, Berlin.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Intrator, O. and Kooperberg, C. (1995). Trees and splines in survival analysis. *Statist. Meth. Med. Res.* **4** 237–261.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995a). Hazard regression. *J. Amer. Statist. Assoc.* **90** 78–94.

Kooperberg, C., Stone, C. J. and Truong, Y. K. (1995b). The L_2 rate of convergence for hazard regression. *Scand. J. Statist.* **22** 143–157.

Kooperberg, C., Stone, C. J. and Truong, Y. K. (1996). Entry on 'Hazard regression'. To appear in the *Encyclopedia of Statistical Sciences.* Wiley.

Schumaker, L. L. (1981). *Spline Functions: Basic Theory.* Wiley, New York.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

Stone, C. J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday* (M. H. Rezvi, J. S. Rustagi and D. Siegmund, eds.) 393–406. Academic Press, New York.

Stone, C. J. (1989). Uniform error bounds involving logspline models. In *Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin* (T. W. Anderson, K. B. Athreya and D. L. Iglehart, eds.) 335–355. Academic Press, Boston.

Stone, C. J. (1990). Large-sample inference for log-spline models. *Ann. Statist.* **18** 717–741.

Stone, C. J. (1991). Asymptotics for doubly flexible logspline response models. *Ann. Statist.* **19** 1832–1854.

Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.* **22** 118–184.

Department of Biostatistics
University of North Carolina
Chapel Hill, NC 27599-7400

Department of Statistics
University of California
Berkeley, CA 94720-3860