Probab. Th. Rel. Fields 87, 1–25 (1990)

Probability Theory and Related Fields

Identifying nonlinear covariate effects in semimartingale regression models

Ian W. McKeague [1,*] **and Klaus J. Utikal** [2,**]

[1] Department of Statistics, The Florida State University, Tallahassee, FL 32306-3033, USA
[2] Department of Statistics, University of Kentucky, Lexington, KY 40506, USA

Received May 13, 1988; in revised form March 19, 1990

Summary. Let X_t be a semimartingale which is either continuous or of counting process type and which satisfies the stochastic differential equation $dX_t = Y_t\,\alpha(t, Z_t)\,dt + dM_t$, where Y and Z are predictable covariate processes, M is a martingale and α is an unknown, nonrandom function. We study inference for α by introducing an estimator for $\mathscr{A}(t,z) = \int_0^z \int_0^t \alpha(s,x)\,ds\,dx$ and deriving a functional central limit theorem for the estimator. The asymptotic distribution turns out to be given by a Gaussian random field that admits a representation as a stochastic integral with respect to a multiparameter Wiener process. This result is used to develop a test for independence of X from the covariate Z, a test for time-homogeneity of α, and a goodness-of-fit test for the proportional hazards model $\alpha(t,z) = \alpha_1(t)\,a_2(z)$ used in survival analysis.

1 Introduction

Consider a nonlinear semimartingale regression model in which a process X is related to a covariate process Z by

$$X_t = X_0 + \int_0^t \lambda_s\,ds + M_t, \tag{1.1}$$

$$\lambda_t = Y_t\,\alpha(t, Z_t), \tag{1.2}$$

where α is an unknown function, M is a martingale and Y is an indicator process, taking the value 1 when X and Z are under observation, zero otherwise. In the case that X is a counting process, λ and α are called the intensity process and conditional hazard function respectively. If the intensity process is of the form $\lambda_t = \alpha(t)\,Z_t$, we have Aalen's (1978) multiplicative intensity model, for which

---

* Research supported by the Army Research Office under Grant DAAL03-86-K-0094
** Research supported by the Air Force Office of Scientific Research under Contract F49620-85-C-0007

a well developed theory of hazard rate and cumulative hazard rate estimation exists (see the survey articles of Andersen and Borgan (1985) and McKeague and Utikal (1990a)). In the classical survival analysis setting, Beran (1981) and Dabrowska (1987) studied an estimator $\hat{A}(\cdot, z)$ of the conditional cumulative hazard function $A(t,z)=\int_0^t \alpha(s,z)\, ds$ for a fixed level z of the covariate Z. McKeague and Utikal (1990b) extended this estimator to the general setting above. The estimator was used to develop methods of inference for the function $\alpha(\cdot, z)$ at fixed z.

In the present paper we study inference for the entire conditional 'hazard' function $\alpha(\cdot,\cdot)$. For that purpose we introduce the estimator

$$\hat{\mathscr{A}}(t,z)=\int_0^z \hat{A}(t,x)\,dx$$

of the *doubly cumulative hazard function*

$$\mathscr{A}(t,z)=\int_0^z \int_0^t \alpha(s,x)\,ds\,dx=\int_0^z A(t,x)\,dx.$$

When X is a continuous process or a counting process we establish the weak convergence of the appropriately normalized time and state indexed process $\hat{\mathscr{A}}$ to a Gaussian random field. This is proved by using the results of Bickel and Wichura (1971) to establish tightness. Convergence of the finite dimensional distributions is shown using Rebolledo's (1980) martingale central limit theorem.

We also propose a test for independence of X from the covariate process Z. Here independence from the covariate means that α is only a function of time. A natural estimator for $\mathscr{A}$ under the hypothesis of independence is given by $\bar{\mathscr{A}}(t,z)=z\,\bar{A}(t)$, where $\bar{A}$ is the Nelson-Aalen estimator. We derive the asymptotic distribution of $\hat{\mathscr{A}}-\bar{\mathscr{A}}$ and show that a maximal deviation statistic based on $\hat{\mathscr{A}}-\bar{\mathscr{A}}$ yields a consistent test for independence.

Furthermore, we propose a test for time-homogeneity, i.e. that $\alpha=\alpha(t,z)$ does not depend on time t. An estimator for $\mathscr{A}$ under the hypothesis of time-homogeneity is given by $\mathscr{A}^*(t,z)=t\,\hat{\mathscr{A}}(1,z)$. A maximal deviation test statistic based on $\hat{\mathscr{A}}-\mathscr{A}^*$ is shown to yield a consistent test for time-homogeneity.

Finally we develop a goodness-of-fit test for the general proportional hazards model $\alpha(t,z)=\alpha_1(t)\,\alpha_2(z)$, where $\alpha_1(t)$ and $\alpha_2(z)$ are arbitrary unknown functions. This model has been studied by Thomas (1983), Tibshirani (1984), Hastie and Tibshirani (1986) and O'Sullivan (1986a, 1986b) in the survival analysis context (where it is a generalization of Cox's (1972) proportional hazards model). These authors propose various estimators for the log relative risk function $\log\alpha_2$, where α_2 is assumed to be positive, but, except for O'Sullivan (1986a), who finds a rate of convergence for his estimator, they do not provide any asymptotic theory. We suggest $\hat{\mathscr{A}}(1,\cdot)$ as an estimator of the cumulative relative risk function $A_2(\cdot)=\int_0^{\cdot} \alpha_2(x)\,dx$ and find its asymptotic distribution.

The precise description of our model and some preliminaries are given in Section 2. The doubly cumulative hazard function estimator $\hat{\mathscr{A}}$ is introduced in Section 3, and in Section 4 we derive the various goodness-of-fit tests based on $\hat{\mathscr{A}}$. Technical lemmas used in the proofs of our main results are collected in Section 5.

2 Preliminaries

Let $(\Omega, \mathscr{F}, P)$ be a complete probability space and let $(\mathscr{F}_t, t\in[0,1])$ be a right-continuous filtration, where $\mathscr{F}_0$ contains all P-null sets in $\mathscr{F}$. All processes in this section are indexed by $t\in[0,1]$. The process $(M_t, \mathscr{F}_t)$ is assumed to be a zero-mean L^4-martingale with sample paths in Skorohod space $D[0,1]$. The quadratic characteristic of M will be denoted by $\langle M\rangle$ and its quadratic variation by $[M]$. The processes Y and Z are assumed to be predictable, with Y an indicator process. For simplicity, Z is supposed to be scalar valued.

The processes X, Y, Z and M are related by (1.1) and (1.2) which can be written in the form

(2.1) $$dX_t = Y_t\,\alpha(t, Z_t)\,dt + dM_t.$$

We assume that α is Lipschitz and

(2.2) $$\langle M\rangle_t = \int_0^t \gamma(t, Z_s)\,Y_s\,ds,$$

where γ is a continuous function. Important examples of our model include counting processes, in which case $\gamma=\alpha$, see Examples 1–4 of McKeague and Utikal (1990b); and diffusion processes, in which case α is the drift, γ is the infinitesimal variance and $Y\equiv 1$.

Estimation of $\mathscr{A}$ is to be carried out over the unit square $[0,1]^2$. For that purpose introduce processes X_i, Y_i, Z_i, M_i, $i=1,\dots,n$ and a filtration $(\mathscr{F}_t^{(n)})$ having the same structure as X, Y, Z, M and $(\mathscr{F}_t)$ above. Further suppose that M_i, $i=1,\dots,n$ are orthogonal $\mathscr{F}_t^{(n)}$ martingales. This is the case, for instance, if $\mathscr{F}_t^{(n)}=\mathscr{F}_{1t}\vee\dots\vee\mathscr{F}_{nt}$, where $\mathscr{F}_{1t},\dots,\mathscr{F}_{nt}$ are independent filtrations and each M_i is an $\mathscr{F}_{it}$ martingale. In the counting process case the M_i are orthogonal $\mathscr{F}_t^{(n)}$ martingales if no two of the counting processes X_i jump simultaneously.

Let $W=(W(t,z), (t,z)\in[0,1]^2)$ be a two-parameter Wiener process, i.e. a Gaussian process with zero mean and $EW(t,z)\,W(t',z')=\min(t,t')\,\min(z,z')$. Let $\int_0^t\int_0^z \psi(s,x)\,dW(s,x)$ denote a continuous version of the Wiener integral of a function $\psi\in L^2([0,1]^2, ds\,dx)$ defined by Ito (1951), Wong and Zakai (1974) and Bass (1988). The estimators and test statistics that we shall introduce have asymptotic distributions which can be represented in terms of stochastic integrals of this type. Let C_2 denote the space of continuous functions on the unit square equipped with the supremum norm $\|\cdot\|$. Let D_2 denote the extension of the space $D[0,1]$ to functions on $[0,1]^2$, as described in Neuhaus (1971).

3 The doubly cumulative hazard function estimator

Our estimator is based on observation of (X_i, Y_i, Z_i), $i=1,\dots,n$. The basic idea is to first estimate $A(\cdot, z)$ by stratifying over the covariate and then integrate with respect to z. The strata are given by $\mathscr{I}_r=[x_{r-1}, x_r)$, $r=1, \ldots, d_n$, where $x_r=r/d_n$ and d_n is an increasing sequence of positive integers. The width of each stratum is $w_n=1/d_n$. Define $\mathscr{I}_z=\mathscr{I}_r$ for $z\in\mathscr{I}_r$. Note that $\mathscr{I}_z$ depends implicitly on n.

Let $X_i(t, z)$ denote the contribution to $X_i(t)$ from stratum $\mathscr{I}_z$:

$$X_i(t, z)=\int_0^t I\{Z_i(s)\in\mathscr{I}_z\}\, dX_i(s),$$

and set $X^{(n)}(s, z)=\sum_{i=1}^n X_i(s, z)$. The number of covariate processes observed to be in stratum $\mathscr{I}_z$ at time s is given by

$$Y^{(n)}(s, z)=\sum_{i=1}^n I\{Z_i(s)\in\mathscr{I}_z\}\, Y_i(s).$$

The following estimator of $A(\cdot, z)$ is an extension of Beran's (1981) estimator of a conditional cumulative hazard function:

$$\hat{A}(t, z)=\int_0^t \frac{X^{(n)}(ds, z)}{Y^{(n)}(s, z)},$$

where $1/0\equiv 0$. The stratum width w_n should tend to zero at a suitable rate as $n\to\infty$. Our doubly cumulative hazard function estimator $\hat{\mathscr{A}}$ is now defined by

$$\hat{\mathscr{A}}(t, z)=\int_0^z \hat{A}(t, x)\, dx.$$

In order to derive the asymptotic distribution of $\hat{\mathscr{A}}$ we shall need the following conditions (in which (s, x) is understood to range over the unit square).

Conditions

(A1) There exists a nonnegative continuous function $\phi(\cdot,\cdot)$ such that

$$\iint_{[0,1]^2} \left|\frac{n w_n}{Y^{(n)}(s, x)}-\phi(s, x)\right| ds\, dx \xrightarrow{P} 0.$$

(A2) $\mathrm{Leb}_2\{(s, x):\, Y^{(n)}(s, x)=0\}=o_P\left(\frac{1}{\sqrt{n}}\right)$.

(A3) $\sup_{s, x, n} E\left[\frac{n\omega_n}{Y^{(n)}(s, x)}\right]^3<\infty$.

Theorem 3.1. *Suppose that Conditions* $(A1)-(A3)$ *hold,* $nw_n^2 \to 0$, $d_n = O(n)$ *and* X *is a counting process or has continuous sample paths. Let* $h = \gamma \cdot \phi$. *Then*

$$\sqrt{n}(\hat{\mathscr{A}} - \mathscr{A}) \xrightarrow{\mathscr{D}} m$$

in D_2 *as* $n \to \infty$, *where*

$$m(t, z) = \int_0^t \int_0^z \sqrt{h(s, x)}\, dW(s, x).$$

Remark. The process m is a continuous Gaussian random field with mean zero and covariance function

$$\operatorname{Cov}(m(t_1, z_1), m(t_2, z_2)) = \int_0^{z_1 \wedge z_2} \int_0^{t_1 \wedge t_2} h(s, x)\, ds\, dx.$$

Conditions (A1)−(A3) are easily checked when (X_i, Y_i, Z_i), $i \geqq 1$ are independent copies of (X, Y, Z). We shall need to assume that the covariate subdistribution function $F(s, x) = P(Z_s \leqq x, Y_s = 1)$ satisfies the following mild condition:

(IID)(i.i.d. case) For each $s \in [0, 1]$, $F(s, \cdot)$ is absolutely continuous on $[0, 1]$ and has density $f(s, \cdot)$ such that $f(\cdot, \cdot)$ is continuous and bounded away from zero.

Proposition 3.2. *Suppose that* (IID) *holds,* $\omega_n \to 0$ *and* $nw_n^\delta \to \infty$ *for some* $\delta > 1$. *Then Conditions* $(A1)-(A3)$ *hold with* $\phi = 1/f$.

From Proposition 3.2 we see that the conclusion of Theorem 3.1 holds under Condition (IID) if $nw_n^2 \to 0$ *and* $nw_n^\delta \to \infty$ for some $\delta > 1$ (equivalently, if $d_n^2/n \to \infty$ and $d_n = o(n^\delta)$ for some $\delta < 1$). The implicit assumption for Theorem 3.1 that α is Lipschitz can be weakened to Lipschitz of order η, $0 < \eta \leqq 1$, if $nw_n^2 \to 0$ is strengthened to $nw_n^{2\eta} \to 0$.

There is a useful example of our model for which Conditions (A1)−(A3) are too restrictive. For the illness-death process with duration dependence, considered in Example 3 of McKeague and Utikal (1990b), the density f vanishes over part of the unit square. In that case (A1)−(A3) can be proved only when (s, x) vary over the region

$$\mathscr{J} = \{(s, z) \in [0, 1]^2 : \mathscr{I}_z \text{ is contained in the support of } \alpha(s, \cdot)\}.$$

It is possible to deal with such cases by extending Theorem 3.1 along the lines of Theorem 1 of McKeague and Utikal (1990b).

It would be preferable to estimate $\mathscr{A}$ without the use of stratification or any other smoothing technique, as it is preferable to estimate a bivariate *cdf* by the bivariate empirical *cdf* rather than estimating one of the conditional *cdf's* and then integrating with respect to an estimated marginal distribution. However, for the general nonlinear semimartingale regression model considered here, we doubt that an efficient non-smoothing based estimator of $\mathscr{A}$ could be constructed. In the counting process case, Greenwood and Wefelmeyer (1989) have recently shown that $\hat{\mathscr{A}}$ is efficient in the sense of satisfying a Hájek-Le Cam convolution theorem. Thus, even if a more direct estimator exists, there is no (asymptotic) disadvantage in using $\hat{\mathscr{A}}$ in that case.

Although $\hat{\mathscr{A}}$ uses smoothing, it is relatively insensitive to the choice of w_n, as suggested by the weak conditions on w_n in Theorem 3.1. The estimator $\hat{A}(t, x)$ is sensitive to w_n, but this effect is largely drowned out following integration over x. Reducing w_n would give only a slightly rougher $\hat{\mathscr{A}}$, increasing w_n would tend to smooth $\hat{\mathscr{A}}$. In practice we would allow the strata $\mathscr{I}_r$ to have unequal widths and choose the intervals adaptively to ensure that the covariate enters each stratum sufficiently often. In numerical experiments we have obtained satisfactory results when using $d_n = 10$ for $n = 100$, and $d_n = 25$ for $n = 500$.

Proof of Theorem 3.1. Define the processes

$$M^{(n)}(t, z) = \sum_{i=1}^{n} \int_0^t I\{Z_i(s) \in \mathscr{I}_z\}\, dM_i(s),$$

$$\alpha^{(n)}(t, z) = \sum_{i=1}^{n} I\{Z_i(t) \in \mathscr{I}_z\}\, Y_i(t)\, \alpha(t, Z_i)(t)),$$

so that by (2.1), $\sqrt{n}(\hat{\mathscr{A}} - \mathscr{A}) = \hat{M} + R$, where

$$\hat{M}(t, z) = \sqrt{n} \int_0^z \int_0^t \frac{M^{(n)}(ds, x)}{Y^{(n)}(s, x)}\, dx,$$

$$R(t, z) = \sqrt{n} \int_0^z \int_0^t \left[\frac{\alpha^{(n)}(s, x)}{Y^{(n)}(s, x)} - \alpha(s, x)\right] ds\, dx.$$

It suffices to show that $\|R\| \xrightarrow{P} 0$ and $\hat{M} \xrightarrow{\mathscr{D}} m$ in D_2. Since α is assumed to be Lipschitz,

$$\|R\| \leq \sqrt{n} \iint_{[0,1]^2} \left| \frac{(\alpha(s, x) + O(w_n))\, Y^{(n)}(s, x)}{Y^{(n)}(s, x)} - \alpha(s, z) \right| ds\, dx$$

$$= O(\sqrt{n})\, \mathrm{Leb}_2\{(s, x)\colon Y^{(n)}(s, x) = 0\} + O(\sqrt{n w_n^2}) \xrightarrow{P} 0,$$

by (A 2) and $n w_n^2 \to 0$. Thus $\|R\| \xrightarrow{P} 0$.

Introduce the following approximation to $\hat{M}$ which is piecewise constant over $z \in \mathscr{I}_r$:

$$\tilde{M}(t, z) = \sqrt{n w_n^2} \sum_{r=1}^{[z d_n]} \int_0^t \frac{dM_r^{(n)}(s)}{Y_r^{(n)}(s)}, \tag{3.1}$$

where $M_r^{(n)}(t) = M^{(n)}(t, x_r)$ and $Y_r^{(n)}(t) = Y^{(n)}(t, x_r)$. Here and in the sequel, any summation over $r = 1, \ldots, [z d_n]$ is defined to be zero when $[z d_n] = 0$.

Suppose that $\tilde{M} \xrightarrow{\mathscr{D}} m$ in D_2. Define a linear map $\pi_n\colon D_2 \to D_2$ by

$$\pi_n(\psi)(t, x) = \psi(t, x_{r-1}) + (\psi(t, x_r) - \psi(t, x_{r-1}))(x - x_{r-1})/w_n$$

for $x \in \mathscr{I}_r$. Here $\pi_n(\psi)(t, \cdot)$ is a piecewise linear approximation to $\psi(t, \cdot)$ based on its values at x_r, $r = 1, \ldots, d_n$, for each t. Note that $\hat{M} = \pi_n(\tilde{M})$. Appealing to a D_2 version of Lemma 4.1 of McKeague (1988), we have $\pi_n(\tilde{M}) \xrightarrow{\mathscr{D}} m$ in D_2, where we have used the fact that m has its sample paths in C_2. Thus $\hat{M} \xrightarrow{\mathscr{D}} m$ in D_2. All that remains to be proved is that $\tilde{M} \xrightarrow{\mathscr{D}} m$ in D_2. This will be established by showing that $\{\tilde{M}, n \geqq 1\}$ is tight in D_2 and the finite dimensional distributions of $\tilde{M}$ converge weakly to those of m.

Denote the increment of $\tilde{M}$ over the rectangle $(s,t] \times (x,y]$ by

$$\tilde{M}((s,t] \times (x,y]) = \tilde{M}(t,y) - \tilde{M}(s,y) - \tilde{M}(t,x) + \tilde{M}(s,x).$$

Tightness is established by checking some product moment conditions of Bickel and Wichura (1971) for the increments of $\tilde{M}$ over certain neighboring rectangles:

$$E(\tilde{M}((s,t] \times (x,y]))^2 (\tilde{M}((s,t] \times (y,z]))^2 \leqq C(t-s)^{\frac{3}{2}}(y-x)(z-y)$$

and

$$E(\tilde{M}((s,t] \times (x,y]))^2 (\tilde{M}((t,u] \times (x,y]))^2 \leqq C(t-s)^{\frac{1}{2}}(u-t)(y-x)^2,$$

where C is a generic positive constant. This is done in Lemmas 2 and 3.

To show convergence of all finite dimensional distributions it suffices to show that for any $0 \leqq z_0 < \ldots < z_p \leqq 1$, $p \geqq 1$

$$(\tilde{M}(\cdot, z_j) - \tilde{M}(\cdot, z_{j-1}))_{j=1}^p \xrightarrow{\mathscr{D}} (m(\cdot z_j) - m(\cdot, z_{j-1}))_{j=1}^p$$

in $D[0,1]^p$, where $D[0,1]^p$ is the product of p copies of $D[0,1]$. This can be done using a p-variate version of Rebolledo's (1980) martingale central limit theorem, as given by Aalen (1977) and Andersen and Gill (1982, Theorem I.2) in the counting process case. The processes $\tilde{M}(\cdot, z_j) - \tilde{M}(\cdot, z_{j-1})$, $j = 1, \ldots, p$ are orthogonal square integrable martingales and by Lemma 4

$$\langle \tilde{M}(\cdot, z_j) - \tilde{M}(\cdot, z_{j-1}) \rangle_t \xrightarrow{P} \langle m(\cdot, z_j) - m(\cdot, z_{j-1}) \rangle_t,$$

for each t, $j = 1, \ldots, p$. That completes the proof for the continuous case. In the counting process case we also need to check the Lindeberg condition (cf. Andersen and Gill's (I.4) with $l = r$, $i = j$, $n = d_n$)

$$\sum_{r=1}^{d_n} \int_0^1 H_{jr}^{(n)}(s)^2 I\{|H_{jr}^{(n)}(s)| > \varepsilon\}\, d\langle M_r^{(n)} \rangle_s \xrightarrow{P} 0, \tag{3.2}$$

for all $\varepsilon > 0$, where

$$H_{jr}^{(n)}(s) = \begin{cases} \sqrt{n w_n^2}\,(Y_r^{(n)}(s))^{-1} & \text{if } [z_{j-1} d_n] < r \leqq [z_j d_n] \\ 0 & \text{otherwise.} \end{cases}$$

This is easily done by noting that $|H_{jr}^{(n)}(s)| \leqq \sqrt{n w_n^2} \to 0$, since $Y_r^{(n)}(s) \geqq 1$, unless it vanishes, so the sum in (3.2) vanishes for sufficiently large n. □

Proof of Proposition 3.2. The result follows from Lemma 1. Part (c) of that lemma gives (A1), and by part (b)

$$E\,\mathrm{Leb}_2\{(s,x)\colon Y^{(n)}(s,x)=0\} = \iint_{[0,1]^2} P(Y^{(n)}(s,x)=0)\,ds\,dx$$
$$\leqq e^{-Knw_n} = o\left(\frac{1}{\sqrt{n}}\right),$$

using $nw_n^\delta \to \infty$ for some $\delta > 1$, giving (A2). Condition (A3) follows from part (a) of the lemma with $k=3$. □

Confidence sets for $\mathscr{A}$

In order to apply Theorem 3.1 to obtain Kolmogorov-Smirnov type confidence sets for $\mathscr{A}$ of the form $\{\mathscr{A}\colon \sqrt{n}\sup_{t,z}|\hat{\mathscr{A}}(t,z)-\mathscr{A}(t,z)|\leqq c\}$ we would need to determine the quantiles of $\sup_{t,z}|m(,z)|$. In the time-homogeneous case, considered below, it is possible to use existing tables. In the general case, the representation of m in terms of the Brownian sheet process W gives a way to obtain such quantiles by simulation. We shall only consider this in the counting process case, but the continuous case is similar. First estimate the function $H(t,z) = \int_0^t \int_0^z h(s,x)\,dx\,ds$ by

$$\hat{H}(t,z) = nw_n \int_0^z \int_0^t \frac{X^{(n)}(ds,x)}{(Y^{(n)}(s,x))^2}\,dx$$

and then estimate h by

$$\hat{h}(t,z) = \frac{1}{b_n^2} \iint_{[0,1]^2} K\left(\frac{t-s}{b_n}\right) K\left(\frac{z-x}{b_n}\right) d\hat{H}(s,x).$$

where K is a bounded, nonnegative kernel function with compact support, integral 1 and b_n is a bandwidth parameter, $b_n \to 0$. The following result, proved at the end of Section 5, shows that $\hat{h}$ is an L^1-consistent estimator of h.

Proposition 3.3. *If X is a counting process, the assumptions of Proposition* 3.2 *hold,* $w_n = o(b_n^2)$, *and K is Lipschitz, then* $E \iint_{[0,1]^2} |\hat{h}(t,z)-h(t,z)|\,dt\,dz \to 0.$

The process m, with $\hat{h}$ in place of h, could then be simulated to obtain approximate quantiles for $\sup_{t,z}|m(t,z)|$. Using Proposition 3.3 it can be shown (see the proof of Proposition 4.2) that this procedure leads to asymptotically correct confidence sets for $\mathscr{A}$.

Time-homogeneous counting process

Let N be a counting process which is time-homogeneous in the sense that its conditional hazard function only depends on the covariate process Z, so N has intensity

$$\lambda_t = Y_t\,\alpha(Z_t).$$

An estimator of $\mathscr{A}(z)=\int_0^z \alpha(x)\,dx$ from i.i.d. copies (N_i, Y_i, Z_i), $i=1,\dots,n$ of (N, Y, Z) is given by

(3.3)
$$\hat{\mathscr{A}}(z)=\int_0^z\int_0^1 \frac{N^{(n)}(ds,x)}{Y^{(n)}(s,x)}\,dx,$$

where $N^{(n)}$ is defined as $X^{(n)}$, but with N in place of X. To apply our result to this special case we note that the projection $\pi\colon D_2\to D[0,1]$ defined by $\pi(\psi)(z)=\psi(1,z)$ is continuous, so by the continuous mapping theorem (e.g. Billingsley, 1968, Theorem 5.1) we obtain the following consequence of Theorem 3.1. A similar result could be obtained in the case that X has continuous sample paths.

Proposition 3.4. *Suppose that Condition* (*IID*) *holds,* $nw_n^2\to 0$ *and* $nw_n^\delta\to\infty$ *for some* $\delta>1$. *Then, for* $\hat{\mathscr{A}}$ *defined by* (3.3),

$$\sqrt{n}(\hat{\mathscr{A}}-\mathscr{A})\xrightarrow{\mathscr{D}} m$$

in $C[0,1]$ *as* $n\to\infty$, *where* $m=(m(z),\ z\in[0,1])$ *is a continuous Gaussian martingale with mean zero and covariance function* $\operatorname{Cov}(m(z_1), m(z_2))=H(z_1\wedge z_2)$, *where*

$$H(z)=\int_0^z\int_0^1 \frac{\alpha(x)}{f(s,x)}\,ds\,dx.$$

With the help of Proposition 3.4 we now construct confidence bands for $\mathscr{A}$. Denote

$$\hat{H}(z)=nw_n\int_0^z\int_0^1 \frac{N^{(n)}(ds,x)}{(Y^{(n)}(s,x))^2}\,dx.$$

As a consequence of the proposition,

$$\sqrt{n}\,\frac{\sqrt{H(1)}}{H(\cdot)+H(1)}(\hat{\mathscr{A}}(\cdot)-\mathscr{A}(\cdot))\xrightarrow{\mathscr{D}} W^0\left(\frac{H(\cdot)}{H(\cdot)+H(1)}\right)$$

in $C[0,1]$ as $n\to\infty$, where W^0 is a standard Brownian bridge. Now $\hat{H}$ is a uniformly consistent estimator of H by Lemma 7. Thus we obtain the following asymptotic $100(1-\beta)\%$ confidence band for $\mathscr{A}$:

$$\hat{\mathscr{A}}(z)\pm c_\beta\sqrt{\frac{\hat{H}(1)}{n}}\left(1+\frac{\hat{H}(z)}{\hat{H}(1)}\right),\qquad z\in[0,1],$$

where $P(\sup_{0\leqq t\leqq 1/2}|W^0(t)|\geqq c_\beta)=\beta$, $0<\beta<1$. A table for c_β can be found in Hall and Wellner (1980).

4 Goodness-of-fit tests

4.1 Testing for independence from the covariate process

In this subsection we consider the problem of testing whether the covariate process Z is absent from the model, i.e. whether α is only a function of time. Such a test could be used to test whether a pure jump process on a finite state space is a Markov process, see McKeague and Utikal (1990b, Example 2). Let H_0 denote the null hypothesis H_0: $\alpha(t,z_1)=\alpha(t,z_2)$ for all t, z_1, $z_2\in[0,1]$. Under H_0 the natural estimator of $\mathscr{A}$ is

$$\bar{\mathscr{A}}(t,z)=z\bar{A}(t),$$

where $\bar{A}$ is the Nelson-Aalen estimator

$$\bar{A}(t)=\int_0^t\frac{d\bar{X}^{(n)}(s)}{\bar{Y}^{(n)}(s)}$$

and

$$\bar{X}^{(n)}(t)=\sum_{i=1}^n\int_0^t I(Z_i(s)\in[0,1])\,dX_i(s),$$

$$\bar{Y}^{(n)}(s)\sum_{i=1}^n I(Z_i(s)\in[0,1])\,Y_i(s).$$

Define some functions g and ρ by

$$g(s,x)=\gamma(s,x)f(s,x)/\rho^2(s),$$

$$\rho(s)=P(Z_s\in[0,1],\,Y_s=1)=\int_0^1 f(s,x)\,dx.$$

The following result gives the asymptotic distribution of $\hat{\mathscr{A}}-\bar{\mathscr{A}}$.

Theorem 4.1. *Suppose that Condition (IID) holds,* $nw_n^2\to 0$ *and* $nw_n^\delta\to\infty$ *for some* $\delta>1$. *Then, under* H_0,

$$\sqrt{n}(\hat{\mathscr{A}}-\bar{\mathscr{A}})\xrightarrow{\mathscr{D}} m_0$$

in D_2 *as* $n \to \infty$, *where*

$$m_0(t,z) = \int_0^t \int_0^z \sqrt{h(s,x)}\, dW(s,x) - z \int_0^t \int_0^1 \sqrt{g(s,x)}\, dW(s,x).$$

The Kolmogorov-Smirnov type statistic $T^{(n)} = \sqrt{n} \sup_{t,z} |\hat{\mathscr{A}}(t,z) - \tilde{\mathscr{A}}(t,z)|$ could be used for testing H_0. Note that the continuous mapping theorem and Theorem 4.1 imply that $T^{(n)} \xrightarrow{\mathscr{D}} \sup_{t,z} |m_0(t,z)|$ as $n \to \infty$. In order to construct an asymptotic size β test of H_0, rejecting H_0 if $T^{(n)}$ is large, we first need to introduce appropriate estimators for the functions g and h under H_0. Again we shall only do this in the counting process case. Let

$$\bar{G}(t,z) = \frac{n}{w_n} \int_0^z \int_0^t \frac{Y^{(n)}(s,x)}{(\bar{Y}^{(n)}(s))^3}\, d\bar{X}^{(n)}(s)\, dx,$$

$$\bar{H}(t,z) = n w_n \int_0^z \int_0^t \frac{d\bar{X}^{(n)}(s)}{Y^{(n)}(s,x)\, \bar{Y}^{(n)}(s)}\, dx$$

and define

$$\bar{g}(t,z) = \frac{1}{b_n^2} \int_0^1 \int_0^1 K\left(\frac{t-s}{b_n}\right) K\left(\frac{z-x}{b_n}\right) d\bar{G}(s,x),$$

$$\bar{h}(t,z) = \frac{1}{b_n^2} \int_0^1 \int_0^1 K\left(\frac{t-s}{b_n}\right) K\left(\frac{z-x}{b_n}\right) d\bar{H}(s,x),$$

where K is a bounded, nonnegative kernel function with compact support, integral 1 and b_n is a bandwidth parameter, $b_n \to 0$.

The distribution of $T = \sup_{t,z} |m_0(t,z)|$ depends only on $\theta = (g, h)$ and is continuous, see Ylvisaker (1968). Let $c_\beta(\theta)$ denote the upper β-quantile of T, so that $P_\theta\{T > c_\beta(\theta)\} = \beta$ for $0 < \beta < 1$. Given the estimate $\hat{\theta}_n = (\bar{g}, \bar{h})$, we may simulate the process m_0, with $\bar{g}$ and $\bar{h}$ in place of g and h respectively, to obtain an approximate critical level $c_\beta^{(n)} = c_\beta(\hat{\theta}_n)$. In Proposition 4.2 we show that

$$\lim_{n \to \infty} P(T^{(n)} > c_\beta^{(n)}) = \beta.$$

Thus, rejecting H_0 when $T^{(n)} > c_\beta^{(n)}$ yields an asymptotic size β test for independence. In Proposition 4.3 we show that this test is consistent against all alternatives.

Proof of Theorem 4.1. Decomposing $\tilde{\mathscr{A}}$ in a similar way to $\hat{\mathscr{A}}$ in the proof of Theorem 3.1, we can write

$$\sqrt{n}(\hat{\mathscr{A}} - \tilde{\mathscr{A}})(t,z) = \hat{M}(t,z) - z\bar{M}(t) + R(t,z) - \bar{R}(t,z),$$

where

$$\bar{M}(t)=\sqrt{n}\int_0^t \frac{d\bar{M}^{(n)}(s)}{\bar{Y}^{(n)}(s)}$$

$$\bar{M}^{(n)}(t)=\sum_{i=1}^n \int_0^t I\{Z_i(s)\in[0,1]\}\, dM_i(s),$$

R is defined in the proof of Theorem 3.1, and under H_0

$$\bar{R}(t,z)=-z\int_0^t \alpha(s)\, I(\bar{Y}^{(n)}(s)=0)\, ds.$$

Putting $w_n\equiv 1$ in Lemma 1 (b), we obtain $\|\bar{R}\|\xrightarrow{P} 0$ under H_0. Also, by the proof of Theorem 3.1, $\|R\|\xrightarrow{P} 0$. To complete the proof it suffices to show that $\xi\xrightarrow{\mathscr{D}} m_0$, where $\xi(t,z)=\hat{M}(t,z)-z\bar{M}(t)$. Set

$$\bar{m}(t)=\int_0^t\int_0^1 \sqrt{g(s,x)}\, dW(s,x),$$

where W is the same Brownian sheet used to define m in Theorem 3.1. Then $\bar{m}$ is a zero mean continuous Gaussian martingale with predictable variation process

$$\langle\bar{m}\rangle_t=\int_0^t\int_0^1 g(s,x)\, dx\, ds.$$

Suppose that $(\tilde{M},\bar{M})\xrightarrow{\mathscr{D}}(m,\bar{m})$ jointly in $D_2\times D[0,1]$. Define a map π'_n: $D_2\times D[0,1]$ by $\pi'_n(\psi_1,\psi_2)=(\pi_n(\psi_1),\psi_2)$, where π_n is defined in the proof of Theorem 3.1. Then, as in that proof, $(\hat{M},\bar{M})=\pi'_n(\tilde{M},\bar{M})\xrightarrow{\mathscr{D}}(m,\bar{m})$ jointly in $D_2\times D[0,1]$ and, since m and $\bar{m}$ have continuous paths, by the continuous mapping theorem we may conclude that ξ converges weakly to the process $m(t,z)-z\bar{m}(t)=m_0(t,z)$.

It remains to show that $(\tilde{M},\bar{M})\xrightarrow{\mathscr{D}}(m,\bar{m})$ jointly in $D_2\times D[0,1]$. The process $\bar{M}$ is a martingale and $\langle\bar{M}\rangle_t\xrightarrow{P}\langle\bar{m}\rangle_t$, by Lemma 5(a). The Lindeberg condition

$$\int_0^1 \bar{H}^{(n)}(s)^2 I\{|\bar{H}^{(n)}(s)|>\varepsilon\}\, d\langle\bar{M}^{(n)}\rangle_s\xrightarrow{P} 0, \tag{4.1}$$

for all $\varepsilon>0$, where $\bar{H}^{(n)}(s)=\sqrt{n}(\bar{Y}^{(n)}(s))^{-1}$, is checked in Lemma 6. Therefore, by Rebolledo's martingale central limit theorem, $\bar{M}\xrightarrow{\mathscr{D}}\bar{m}$, in $D[0,1]$. Also, by the proof of Theorem 3.1, we have $\tilde{M}\xrightarrow{\mathscr{D}} m$ in D_2. If we can show that the finite dimensional distributions of $(\tilde{M},\bar{M})$ converge to those of $(m,\bar{m})$, then $(\tilde{M},\bar{M})\xrightarrow{\mathscr{D}}(m,\bar{m})$ jointly in $D_2\times D[0,1]$.

To show that the finite dimensional distributions of $(\tilde{M}, \bar{M})$ converge to those of $(m, \bar{m})$, it suffices to show that for any $0 \leqq z_0 < z_1 < \ldots < z_p \leqq 1$, $p \geqq 1$,

$$((\tilde{M}(\cdot, z_j) - \tilde{M}(\cdot, z_{j-1}))_{j=1}^p, \bar{M}(\cdot)) \xrightarrow{\mathscr{D}} ((m(\cdot, z_j) - m(\cdot, z_{j-1}))_{j=1}^p, \bar{m}(\cdot))$$

in $D[0,1]^{p+1}$. This is done using Rebolledo's martingale central limit theorem, as in the proof of Theorem 3.1. It only remains to consider the covariation between $\tilde{M}(\cdot, z)$ and $\bar{M}(\cdot)$. By Lemma 5(b)

$$\langle \tilde{M}(\cdot, z), \bar{M}(\cdot) \rangle_t \xrightarrow{P} \langle m(\cdot, z), \bar{m}(\cdot) \rangle_t,$$

for each z. There are $p+1$ Lindeberg conditions to check. But these conditions have already been checked separately for the p components involving $\tilde{M}$ and the one component involving $\bar{M}$. □

Proposition 4.2. *Suppose that X is a counting process, the assumptions of Theorem 4.1. hold, $d_n b_n^2 \to \infty$ and K is Lipschitz. Then if H_0 holds, for all $0 < \beta < 1$*

$$\lim_{n \to \infty} P(T^{(n)} > c_\beta^{(n)}) = \beta.$$

Proof. Let Θ denote the space of all functions of the form $\theta = (g, h)$ with g and h nonnegative bounded functions on $[0,1]^2$, and endow it with the product metric from $L^1([0,1]^2, ds\,dx) \times L^1([0,1]^2, ds\,dx)$. Let $\theta_n = (g_n, h_n)$, $n \geqq 1$ be a sequence in Θ such that $\theta_n \to \theta$. Then $\sqrt{g_n} \to \sqrt{g}$ and $\sqrt{h_n} \to \sqrt{h}$ in $L^2([0,1]^2, ds\,dx)$. An argument using Doob's inequality applied twice (cf. Cairoli (1970) and Bass (1988)) shows that if $\psi \in L^2([0,1]^2, ds\,dx)$, then

$$E \sup_{t,z} \left| \int_0^t \int_0^z \psi(s,x)\, dW(s,x) \right|^2 \leqq 16 \int_0^1 \int_0^1 \psi^2(s,x)\, ds\, dx.$$

Applying this inequality to $\psi = \sqrt{g_n} - \sqrt{g}$ and $\psi = \sqrt{h_n} - \sqrt{h}$ gives $F_{\theta_n} \xrightarrow{\mathscr{D}} F_\theta$, where F_θ is the distribution function of T under P_θ. Let F_θ^{-1} denote the left-continuous inverse of F_θ. By Billingsley (1986, p. 343) we get $c_\beta(\theta_n) = F_{\theta_n}^{-1}(1-\beta) \to F_\theta^{-1}(1-\beta) = c_\beta(\theta)$, provided F_θ^{-1} is continuous at $1-\beta$. Now, by a version of Proposition 3.3 for $\bar{g}$ and $\bar{h}$, we have $\hat{\theta}_n \xrightarrow{P_\theta} \theta$ in the metric of Θ. Thus, using a subsequence argument, $c_\beta(\hat{\theta}_n) \xrightarrow{P_\theta} c_\beta(\theta)$ and

$$P_\theta(T^{(n)} > c_\beta(\hat{\theta}_n)) \to P_\theta(T > c_\beta(\theta)) = \beta, \tag{4.2}$$

for all but countably many β, where we have used Slutsky's theorem and the continuity of F_θ. Since $P_\theta(T^{(n)} > c_\beta(\hat{\theta}_n))$ is a nondecreasing function of β it follows that (4.2) holds for all $0 < \beta < 1$. □

Proposition 4.3. *Under the assumptions of Theorem 4.1, if H_0 does not hold then $T^{(n)} \xrightarrow{P} \infty$ as $n \to \infty$.*

Proof. First note that if H_0 does not hold then $\|\mathscr{A}-\mathscr{A}_0\|>0$, where

$$\mathscr{A}_0(t,z)=z\int_0^t \frac{1}{\rho(s)}\int_0^1 \alpha(s,x) f(s,x)\,dx\,ds.$$

Using similar arguments to the proof of Lemma 5

$$E\|\bar{\mathscr{A}}_p-\mathscr{A}_0\|^2 \leqq \sup_s E\left|\frac{\bar{\alpha}^{(n)}(s)}{\bar{Y}^{(n)}(s)}-\frac{1}{\rho(s)}\int_0^1 \alpha(s,x) f(s,x)\,dx\right|^2 = O\left(\frac{1}{n}\right),$$

where

$$\bar{\mathscr{A}}_p(t,z)=z\int_0^t \frac{\bar{\alpha}^{(n)}(s)}{\bar{Y}^{(n)}(s)}\,ds \quad \text{and} \quad \bar{\alpha}^{(n)}(s)=\sum_{i=1}^n I\{Z_i(s)\in[0,1]\}\, Y_i(s)\,\alpha(s,Z_i(s)).$$

Also note that by Doob's inequality and Lemma 5(a) we have $E\|\bar{\mathscr{A}}-\bar{\mathscr{A}}_p\|^2 = E\|\bar{M}\|^2/n=O(1/n)$. Thus, since $\sqrt{n}\|\mathscr{A}-\hat{\mathscr{A}}\|=O_P(1)$ by Theorem 3.1,

$$\begin{aligned}\sqrt{n}\|\mathscr{A}-\mathscr{A}_0\| &\leqq \sqrt{n}\|\mathscr{A}-\hat{\mathscr{A}}\|+T^{(n)}+\sqrt{n}\|\bar{\mathscr{A}}-\bar{\mathscr{A}}_p\|+\sqrt{n}\|\bar{\mathscr{A}}_p-\mathscr{A}_0\|\\ &=T^{(n)}+O_P(1).\end{aligned}$$

This shows that $T^{(n)} \xrightarrow{P} \infty$ if H_0 does not hold. □

Remark. The above test for independence can be modified to provide a goodness-of-fit test for Aalen's multiplicative intensity model. Now H_0 is the null hypothesis H_0: there exists a function α_0 such that $\alpha(t,z)=\alpha_0(t)\,z$ for all $t, z\in[0,1]$. Under this H_0, the natural estimator of $\mathscr{A}$ is $\bar{\mathscr{A}}(t,z)=\frac{1}{2}z^2\bar{A}(t)$, where $\bar{A}$ is the Nelson-Aalen estimator as before, except that

$$\bar{Y}^{(n)}(s)=\sum_{i=1}^n I(Z_i(s)\in[0,1])\,Y_i(s)\,Z_i(s).$$

The only changes to Theorem 4.1 are that $\rho(s)=\int_0^1 x f(s,x)\,dx$ and

$$m_0(t,z)=\int_0^t\int_0^z \sqrt{h(s,x)}\,dW(s,x)-\tfrac{1}{2}z^2\int_0^t\int_0^1 \sqrt{g(s,x)}\,dW(s,x).$$

An alternative test of H_0 can be obtained via Theorem 4.1 by partitioning the time × covariate space into cells and forming a chi-squared statistic based on the increments of $\hat{\mathscr{A}}-\bar{\mathscr{A}}$ over the cells, see McKeague and Utikal (1988). This approach is more tractable, but would result in a loss of power. We plan to study the practical aspects of carrying out such tests in a future paper.

4.2 Testing for time-homogeneity

In this subsection we derive a test for the hypothesis $H_0: \alpha(t_1, z) = \alpha(t_2, z)$ for all $t_1, t_2, z \in [0, 1]$, i.e. α is only a function of the covariate. One possible application of this test would be in testing whether a pure jump process on a finite state space is a semi-Markov or Markov renewal process, see McKeague and Utikal (1990b, Example 2). The natural estimator for $\mathscr{A}$ under H_0 is $\mathscr{A}^*(t, z) = t\hat{\mathscr{A}}(1, z)$. In order to test H_0 we could use the test statistic $S^{(n)} = \sqrt{n} \sup_{t,z} |\hat{\mathscr{A}}(t, z) - \mathscr{A}^*(t, z)|$. As in Section 4.1, once we know the asymptotic distribution of $\sqrt{n}(\hat{\mathscr{A}} - \mathscr{A}^*)$ we can derive an asymptotic size α test for H_0 based on $S^{(n)}$. This test can be shown to be consistent using a proof similar to that of Proposition 4.3. The asymptotic distribution of $\sqrt{n}(\hat{\mathscr{A}} - \mathscr{A}^*)$ is given by the following theorem.

Theorem 4.4. *Under the conditions of Theorem* 3.1, *if* H_0 *holds then*

$$\sqrt{n}(\hat{\mathscr{A}} - \mathscr{A}^*) \xrightarrow{\mathscr{D}} m_1$$

in D_2 *as* $n \to \infty$, *where*

$$m_1(t, z) = \int_0^t \int_0^z \sqrt{h(s, x)}\, dW(s, x) - t \int_0^1 \int_0^z \sqrt{h(s, x)}\, dW(s, x).$$

Proof. Note that $\sqrt{n}(\hat{\mathscr{A}} - \mathscr{A}^*) = \pi(\sqrt{n}(\hat{\mathscr{A}} - \mathscr{A}))$, where $\pi\colon D_2 \to D_2$ defined by $\pi(f)(t, z) = f(t, z) - t f(1, z)$ is continuous. The result follows immediately, using Theorem 3.1 and the continuous mapping theorem. □

4.3 Testing for proportionality

Thomas (1983) introduced the general proportional hazards model $\alpha(t, z) = \alpha_1(t)\,\alpha_2(z)$ in the survival analysis context, where α_1 and α_2 are unknown functions. This model is a generalization of Cox's proportional hazards model to allow for arbitrary covariate dependence while keeping the proportional hazards form. In this subsection we introduce a goodness-of-fit test for the general proportional hazards model. Note that this is not the same as a goodness-of-fit test for Cox's proportional hazards model. However, Cox's model can be treated in a similar fashion, see McKeague and Utikal (1988).

Let H_0 denote the null hypothesis H_0: there exist functions α_1 and α_2 such that $\alpha(t, z) = \alpha_1(t)\,\alpha_2(z)$ for all $t, z \in [0, 1]$. In order that α_1 and α_2 be identifiable we impose the condition $A_1(1) = 1$ under H_0, where $A_1(t) = \int_0^t \alpha_1(s)\, ds$. Equivalently, we could impose the condition $A_2(1) = 1$, where $A_2(z) = \int_0^z \alpha_2(x)\, dx$. A reasonable estimator for $\mathscr{A}$ under H_0 is

$$\mathscr{A}^\dagger(t, z) = \hat{A}_1(t)\, \hat{A}_2(z),$$

where

$$\hat{A}_1(t)=\frac{\hat{\mathscr{A}}(t,1)}{\hat{\mathscr{A}}(1,1)}(\text{with } 1/0\equiv 0) \quad \text{and} \quad \hat{A}_2(z)=\hat{\mathscr{A}}(1,z).$$

In order to test H_0 we could use the test statistic $U^{(n)}=\sqrt{n}\sup_{t,z}|\hat{\mathscr{A}}(t,z)-\mathscr{A}^{\dagger}(t,z)|$. As before, once we know the asymptotic distribution of $\sqrt{n}(\hat{\mathscr{A}}-\mathscr{A}^{\dagger})$, we can derive a test for H_0 based on $U^{(n)}$. This test is consistent against any alternative.

Theorem 4.5. *Under the conditions of Theorem* 3.1, *if* H_0 *holds and* $A_2(1)\neq 0$, *then*

$$\sqrt{n}(\hat{\mathscr{A}}-\mathscr{A}^{\dagger})\xrightarrow{\mathscr{D}} m_2$$

in D_2 *as* $n\to\infty$, *where*

$$m_2(t,z)=\int_0^t\int_0^z\sqrt{h(s,x)}\,dW(s,x)-\eta A_2(z)\int_0^t\int_0^1\sqrt{h(s,x)}\,dW(s,x)$$
$$-\eta A_1(t)\int_0^1\int_0^z\sqrt{h(s,x)}\,dW(s,x)+\eta A_1(t)\,A_2(z)\int_0^1\int_0^1\sqrt{h(s,x)}\,dW(s,x)$$

and $\eta=1/A_2(1)$.

Proof. The result follows readily from Theorem 3.1, using the continuous mapping theorem (cf. the proof of Theor. 4.4) and the identities

$$\begin{aligned}(\hat{\mathscr{A}}-\mathscr{A}^{\dagger})(t,z)&=(\hat{\mathscr{A}}-\mathscr{A})(t,z)-\frac{1}{\mathscr{A}(1,1)}[\hat{\mathscr{A}}(t,1)\,\hat{\mathscr{A}}(1,z)-\mathscr{A}(t,1)\,\mathscr{A}(1,z)]\\&\quad+\hat{A}(t,1)\,\hat{\mathscr{A}}(1,z)\left[\frac{1}{\mathscr{A}(1,1)}-\frac{1}{\hat{\mathscr{A}}(1,1)}\right]\\&=(\hat{\mathscr{A}}-\mathscr{A})(t,z)-\frac{1}{\mathscr{A}(1,1)}[(\hat{\mathscr{A}}(t,1)-\mathscr{A}(t,1))(\hat{\mathscr{A}}(1,z)-\mathscr{A}(1,z))\\&\quad+\mathscr{A}(1,z)(\hat{\mathscr{A}}(t,1)-\mathscr{A}(t,1))+\mathscr{A}(t,1)(\hat{\mathscr{A}}(1,z)-\mathscr{A}(1,z)]\\&\quad+\frac{\mathscr{A}(t,1)\,\mathscr{A}(1,z)}{(\mathscr{A}(1,1))^2}(\hat{\mathscr{A}}(1,1)-\mathscr{A}(1,1))\\&\quad+\frac{\hat{\mathscr{A}}(1,1)-\mathscr{A}(1,1)}{\mathscr{A}(1,1)}\left[\frac{\hat{\mathscr{A}}(t,1)\,\hat{\mathscr{A}}(1,z)}{\hat{\mathscr{A}}(1,1)}-\frac{\mathscr{A}(t,1)\,\mathscr{A}(1,z)}{\mathscr{A}(1,1)}\right].\end{aligned}$$ □

Remark. Under H_0, we have from Theorem 3.1 that $\hat{A}_1$ and $\hat{A}_2$ are uniformly consistent estimators of A_1 and A_2, respectively, and

$$\sqrt{n}(\hat{A}_2-A_2)\xrightarrow{\mathscr{D}} m_3$$

in $D[0,1]$ as $n \to \infty$, where m_3 is a continuous Gaussian martingale with covariance function

$$\operatorname{Cov}(m_3(z_1), m_3(z_2)) = \int_0^{z_1 \wedge z_2} \int_0^1 h(s,x)\,ds\,dx.$$

This could be used to obtain confidence bands for A_2 under the general proportional hazards model, by transforming m_3 to Brownian bridge (cf. the discussion following Prop. 3.2). An analogous result can be obtained for $\hat{A}_1$.

5 Technical lemmas

The following lemma is needed to check Conditions (A1)–(A3) in the i.i.d. case.

Lemma 1. *If Condition (IID) holds, then (a) For each positive integer k,*

$$\sup_{s,x,n} E\left[\frac{n w_n}{Y^{(n)}(s,x)}\right]^k < \infty.$$

(b) *Let $C>0$ be a lower bound for f on $[0,1]^2$. Then*

$$\sup_{s,x} P(Y^{(n)}(s,x)=0) \leqq e^{-Cnw_n}.$$

(c) *Let $\phi = 1/f$. If $n w_n \to \infty$, then*

$$\sup_{s,x} E\left|\frac{n w_n}{Y^{(n)}(s,x)} - \phi(s,x)\right| \to 0.$$

Proof. The proof is based on the fact that, in the i.i.d. case, $Y^{(n)}(s,x)$ has a binomial distribution with parameters n and $\int_{\mathcal{I}_x} f(s,u)\,du$. We refer to Lemma 2 of McKeague and Utikal (1990b) for further details. □

Proof of tightness

Tightness of $\{\tilde{M}, n \geqq 1\}$ in D_2 will be shown by establishing a product moment condition on the increments of $\tilde{M}$ over the grid $T^{(n)} = [0,1] \times \{x_r\colon r=0,\ldots,d_n\}$, where $x_r = r w_n$.

Note that

$$\langle M_r^{(n)} \rangle_t = \sum_{i=1}^n \int_0^t I\{Z_i(s) \in \mathcal{I}_r\}\, \gamma(s, Z_i(s))\, Y_i(s)\, ds \tag{5.1}$$

and, since $M_r^{(n)}$, $r=1,\ldots,d_n$ are orthogonal martingales,

$$\langle \tilde{M}(\cdot, z)\rangle_t = n w_n^2 \sum_{r=1}^{[zd_n]} \int_0^t \frac{d\langle M_r^{(n)}\rangle_s}{(Y_r^{(n)}(s))^2}. \tag{5.2}$$

For fixed $0 \leqq s \leqq 1$, $0 \leqq x < y \leqq 1$, define the martingale

$$M_1(t) = \tilde{M}((s,t] \times (x,y]), \qquad t \geqq s,$$

and denote $m_1 = M_1^2 - \langle M_1 \rangle$.

Lemma 2. *Suppose that Condition (A3) holds, γ is bounded, $d_n = O(n)$, and X is either a continuous process or a counting process. Then there exists a positive constant C such that for all $n \geqq 1$, (s,x), $(t,y) \in T^{(n)}$*

$$E\langle M_1\rangle_t^2 \leqq C(t-s)^2(y-x)^2, \tag{5.3}$$

$$E m_1^2(t) \leqq C(t-s)(y-x)^2. \tag{5.4}$$

Proof. In the sequel C is a generic positive constant which is independent of n and s. Note that, by (5.1) and boundedness of γ, we have $d\langle M_r^{(n)}\rangle_s/ds \leqq C Y_r^{(n)}(s)$. Thus, by (5.2) and Fubini's theorem,

$$\begin{aligned} E\langle M_1\rangle_t^2 &= n^2 w_n^4 \sum_{r,l=[xd_n]+1}^{[yd_n]} E \int_s^t \frac{d\langle M_r^{(n)}\rangle_u}{(Y_r^{(n)}(u))^2} \int_s^t \frac{d\langle M_l^{(n)}\rangle_v}{(Y_l^{(n)}(v))^2} \\ &\leqq n^2 w_n^2 (y-x)^2 C^2 \sup_{r,l} \int_s^t \int_s^t E \frac{1}{Y_r^{(n)}(u)\, Y_l^{(n)}(v)}\, du\, dv. \end{aligned} \tag{5.5}$$

But by Condition (A3)

$$\sup_{r,s} E\left(\frac{1}{Y_r^{(n)}(s)}\right)^2 = O\left(\frac{1}{n w_n}\right)^2. \tag{5.6}$$

Hence, applying the Cauchy-Schwarz inequality to the integrand on the r.h.s. of (5.5), we obtain (5.3).

Now we turn to the proof of (5.4). First we need to obtain an explicit expression for m_1. Integration by parts gives

$$M_1^2(t) = 2\int_0^t M_1(v-)\, dM_1(v) + [M_1]_t.$$

In the case that X has continuous sample paths $[M_1] = \langle M_1\rangle$. In the counting process case

$$[M_1]_t = \sum_{s<v\leqq t} (\Delta M_1(v))^2,$$

where $\Delta M_1(v)=M_1(v)-M_1(v-)$ is the jump in M_1 at time v, so $[M_1]=\langle M_1\rangle+\eta$ where η is the martingale

$$\eta_t=n w_n^2 \sum_{r=[x d_n]+1}^{[y d_n]} \int_s^t \frac{d M_r^{(n)}(v)}{(Y_r^{(n)}(v))^2}.$$

Thus

$$m_1(t)=2 \int_s^t M_1(v-)\, d M_1(v)+\eta_t,$$

where in the continuous sample path case η_t is zero. From this expression we get

$$Em_1^2(t) \leqq 8E \int_s^t M_1^2(v-)\, d\langle M_1\rangle_v+2E\langle\eta\rangle_t. \tag{5.7}$$

In order to obtain an upper bound on the first term on the r.h.s. of (5.7) we shall use the Burkholder-Davis-Gundy inequality (see Dellacherie and Meyer, 1982, p. 287)

$$E \sup_{v\in[s,t]} M_1^4(v) \leqq C\, E[M_1]_t^2. \tag{5.8}$$

By orthogonality of the martingales $M_r^{(n)}$, $r=1, \ldots, d_n$,

$$\begin{aligned}
E\int_s^t M_1^2(v-)\, d\langle M_1\rangle_v &= n w_n^2 \sum_{r=[x d_n]+1}^{[y d_n]} E \int_s^t \frac{M_1^2(v-)}{(Y_r^{(n)}(v))^2}\, d\langle M_r^{(n)}\rangle_v \\
&\leqq n w_n^2\, d_n (y-x)(t-s)\, C \sup_{r,v} E\left(\frac{M_1^2(v-)}{Y_r^{(n)}(v)}\right) \\
&\leqq n w_n\, C(y-x)(t-s)(E[M_1]_t^2)^{\frac{1}{2}}\left(\sup_{r,v} E\left(\frac{1}{Y_r^{(n)}(v)}\right)^2\right)^{\frac{1}{2}} \\
&\quad \text{(by (5.8) and the Cauchy – Schwarz inequality)} \\
&\leqq C(y-x)(t-s)(E[M_1]_t^2)^{\frac{1}{2}},
\end{aligned} \tag{5.9}$$

by (5.6). Now $[M_1]=\langle M_1\rangle+\eta$, so

$$E[M_1]_t^2 \leqq 2E\langle M_1\rangle_t^2+2E\langle\eta\rangle_t \leqq C(t-s)^2(y-x)^2+2E\langle\eta\rangle_t \tag{5.10}$$

by (5.3). Also

$$\begin{aligned}
E\langle\eta\rangle_t &= n^2 w_n^4 \sum_{r=[x d_n]+1}^{[y d_n]} E\int_s^t \frac{d\langle M_r^{(n)}\rangle_v}{(Y_r^{(n)}(v))^4} \\
&\leqq n^2 w_n^4\, d_n (y-x)(t-s)\, C \sup_{r,v} E\left(\frac{1}{Y_r^{(n)}(v)}\right)^3 \\
&\leqq C \frac{d_n}{n} \frac{(y-x)}{d_n}(t-s) \qquad \text{(by Condition (A3))} \\
&\leqq C(y-x)^2(t-s)
\end{aligned} \tag{5.11}$$

since $d_n = O(n)$ and $y - x \geqq 1/d_n$ if $x \neq y$. The desired inequality is now obtained directly from (5.7), (5.9)–(5.11). □

Lemma 3. (Tightness). *Suppose that Condition (A3) holds, γ is bounded, $d_n = O(n)$ and X is either a continous process or a counting process. Then $\{\tilde{M}, n \geqq 1\}$ is tight in D_2.*

Proof. Consider the following increments of $\tilde{M}$ over neighboring rectangles in $[0, 1]^2$. Define M_1 as before and

$$M_2(t) = \tilde{M}((s, t] \times (y, z]),$$
$$M_3(u) = \tilde{M}((t, u] \times (x, y]),$$

where $0 \leqq s < t < u \leqq 1$, $0 \leqq x < y < z \leqq 1$. Suppose that the corner points of the rectangles belong to $T^{(n)}$. Also, denote $m_i = M_i^2 - \langle M_i \rangle$, $i = 1, 2, 3$. From the representation of m_1 in the proof of Lemma 2 it can be seen that m_1 and m_2 are orthogonal martingales. Thus, using the Cauchy-Schwarz inequality and Lemma 2, we get

$$\begin{aligned} EM_1^2(t)\, M_2^2(t) &= E\langle M_1\rangle_t \langle M_2\rangle_t + E m_1(t) \langle M_2\rangle_t + E m_2(t) \langle M_1\rangle_t \\ &\quad + E m_1(t)\, m_2(t) \\ &\leqq C(t-s)^{\frac{3}{2}}(y-x)(z-y). \end{aligned} \tag{5.12}$$

Next, by the martingale property of m_3, we have

$$\begin{aligned} EM_1^2(t)\, M_3^2(u) &= E(M_1^2(t)\, E(M_3^2(u) \mid \mathscr{F}_t^{(n)})) = E(M_1^2(t) \langle M_3\rangle_u) \\ &= E m_1(t) \langle M_3\rangle_u + E\langle M_1\rangle_t \langle M_3\rangle_u, \end{aligned}$$

so that, again using the Cauchy-Schwarz inequality and Lemma 2, we obtain

$$EM_1^2(t)\, M_3^2(u) \leqq C(y-x)^2 (t-s)^{\frac{1}{2}} (u-t). \tag{5.13}$$

The inequalities (5.12) and (5.13) imply that "condition (β, γ)" of Bickel and Wichura (1971, p. 1658) is satisfied with $\beta = 3/2$, $\gamma = 4$, for rectangles whose corner points lie in $T^{(n)}$. Clearly $T^{(n)}$ becomes dense in $[0, 1]^2$ as n grows large. Moreover, $\tilde{M}(t, z)$ is constant as a function of z over each interval $\mathscr{I}_r$, $r = 1, \ldots, d_n$, so the modulus of continuity $\omega_\delta''(\tilde{M})$ defined in Bickel and Wichura can be computed using $T^{(n)}$ instead of $[0, 1]^2$. Tightness of $\{\tilde{M}, n \geqq 1\}$ now follows from the remarks following Theorem 3 of Bickel and Wichura (1971, p. 1665). □

Convergence of finite dimensional distributions

Recall the notation $H(t, z) = \int_0^t \int_0^z h(s, x)\, dx\, ds$.

Lemma 4. *Suppose that Conditions (A1) holds and γ is continuous. Then*

$$\sup_{t, z} |\langle \tilde{M}(\cdot, z)\rangle_t - H(t, z)| \xrightarrow{P} 0.$$

Proof. By (5.1) and continuity of γ,

$$\langle M_r^{(n)}\rangle_t = \int_0^t Y_r^{(n)}(s)(\gamma(s, x_r) + o(1))\, ds$$

uniformly in r and t. Therefore by (5.2), continuity of $h = \gamma \cdot \phi$, and (A1),

$$\sup_{t,z} |\langle \tilde{M}(\cdot, z)\rangle_t - H(t,z)| \leqq \frac{1}{d_n} \sum_{r=1}^{d_n} \int_0^1 \left| \frac{n w_n}{Y_r^{(n)}(s)} (\gamma(s, x_r) + o(1)) - h(s, x_r) \right| ds + o(1)$$

$$= O(1) \iint_{[0,1]^2} \left| \frac{n w_n}{Y^{(n)}(s,z)} - \phi(s,z) \right| ds\, dx + o(1) \xrightarrow{P} 0. \quad \square$$

Lemma 5. *Suppose that Condition (IID) holds, $w_n \to 0$ and $n w_n \to \infty$. Then*
(a) $\langle \bar{M}\rangle_t \xrightarrow{P} \int_0^t \int_0^1 g(s,x)\, dx\, ds$;

(b) $\langle \tilde{M}(\cdot, z), \bar{M}(\cdot)\rangle_t \xrightarrow{P} \int_0^t \frac{1}{\rho(s)} \int_0^z \gamma(s, x, 1)\, dx\, ds.$

Proof. From (2.2)

$$\langle \bar{M}\rangle_t = n \int_0^t \frac{\bar{\gamma}^{(n)}(s)}{(\bar{Y}^{(n)}(s))^2}\, ds, \tag{5.14}$$

where

$$\bar{\gamma}^{(n)}(s) = \sum_{i=1}^n I\{Z_i(s) \in [0,1]\}\, Y_i(s)\, \gamma(s, Z_i(s)).$$

Note that $\bar{\gamma}^{(n)}(s)$ and $\bar{Y}^{(n)}(s)$ are sums of uniformly bounded i.i.d. r.v.'s that have expectations $\int_0^1 \gamma(s,x) f(s,x)\, dx$ and $\rho(s)$ respectively. By the Cauchy-Schwarz inquality and Lemma 1 (a) (with $w_n = 1$ and $k = 4$)

$$E\left| n \frac{\bar{\gamma}^{(n)}(s)}{(\bar{Y}^{(n)}(s))^2} - \int_0^1 g(s,x)\, dx \right|$$

$$\leqq \left\{ E\left[\frac{n}{\bar{Y}^{(n)}(s)}\right]^4 \right\}^{1/2} \left\{ E\left[\frac{1}{n}\bar{\gamma}^{(n)}(s) - \left(\frac{\bar{Y}^{(n)}(s)}{n\rho(s)}\right)^2 \int_0^1 \gamma(s,x) f(s,x)\, dx\right]^2 \right\}^{1/2}$$

$$\leqq C\left\{ \operatorname{Var}(\bar{\gamma}^{(n)}(s)/n) + E\left[\rho(s) - \frac{\bar{Y}^{(n)}(s)}{n}\right]^4 \right\}^{1/2}$$

$$\leqq C\{O(1/n) + O(1/n^2)\}^{1/2},$$

uniformly in s. This proves (a). The key step in the proof of (b) is to notice that $\bar{M}$ can be decomposed as

$$\bar{M}(t)=\sqrt{n}\sum_{r=1}^{d_n}\int_0^t \frac{dM_r^{(n)}(s)}{\bar{Y}^{(n)}(s)}.$$

Then, by (3.1) and orthogonality of the martingales $M_r^{(n)}$, $r=1, \ldots, d_n$, we obtain

$$\langle \tilde{M}(\cdot, z), \bar{M}(\cdot)\rangle_t = n w_n \sum_{r=1}^{[z d_n]}\int_0^t \frac{\gamma_r^{(n)}(s)}{Y_r^{(n)}(s)\,\bar{Y}^{(n)}(s)}\,ds,$$

where

$$\gamma_r^{(n)}(s)=\sum_{i=1}^{n} I\{Z_i(s)\in \mathscr{I}_r\}\, Y_i(s)\,\gamma(s, Z_i(s)).$$

Next, using the same approach as in (a),

$$E\left|\frac{n w_n}{\bar{Y}^{(n)}(s)}\sum_{r=1}^{[z d_n]}\frac{\gamma_r^{(n)}(s)}{Y_r^{(n)}(s)}-\frac{1}{\rho(s)}\int_0^z \gamma(s, x)\,dx\right| \leqq C\{L(s, z)+\operatorname{Var}(\bar{Y}^{(n)}(s)/n)\}^{1/2},$$

where

$$L(s, z)=E\left[\frac{1}{d_n}\sum_{r=1}^{[z d_n]}\frac{\gamma_r^{(n)}(s)}{Y_r^{(n)}(s)}-\int_0^z \gamma(s, x)\,dx\right]^2.$$

The variance term is of order $O(1/n)$ uniformly in s. To deal with $L(s, z)$, use continuity of γ to give $\gamma_r^{(n)}(s)=Y_r^{(n)}(s)(\gamma(s, x_r)+o(1))$ uniformly in r and s. Then, by Lemma 1 (b) and $n w_n \to \infty$,

$$\sup_{s, z} L(s, z) \leqq \sup_{s, r} P(Y_r^{(n)}(s)=0)+o(1) \leqq e^{-C n w_n}+o(1) \to 0,$$

completing the proof of (b). □

Lemma 6. *The Lindeberg condition* (4.1) *holds under Condition* (*IID*).

Proof. By (5.14) and boundedness of γ we have $d\langle \bar{M}^{(n)}\rangle_s/ds \leqq C\bar{Y}^{(n)}(s)$. Thus, it suffices to show that

$$E\int_0^1 \frac{n}{\bar{Y}^{(n)}(s)}\, I\left\{\left|\frac{\sqrt{n}}{\bar{Y}^{(n)}(s)}\right|>\varepsilon\right\} ds \to 0.$$

By the Cauchy-Schwarz inequality, an upper bound for this expression is

$$\sup_s \left\{E\left(\frac{n}{\bar{Y}^{(n)}(s)}\right)^2 P\left(\frac{\sqrt{n}}{\bar{Y}^{(n)}(s)}>\varepsilon\right)\right\}^{1/2}$$

which is of order $O(n^{-1/2})$ by Chebychev's inequality and Lemma 1 (a) (with $w_n=1$ and $k=2$). □

Estimation of H and h

Lemma 7. *Suppose that X is a counting process, Conditions (A1) and (A3) hold and* $n w_n \to \infty$. *Then* $\hat{H}$ *is a uniformly consistent estimator of H, i.e.* $\|\hat{H}-H\| \xrightarrow{P} 0$.

Proof. From (2.1), (5.1) and (5.2) we have

$$\hat{H}(t, z) = \langle \tilde{M}(\cdot, z)\rangle_t + n w_n \int_0^z \int_0^t \frac{M^{(n)}(ds, x)}{(Y^{(n)}(s, x))^2}\, dx.$$

In Lemma 4 we showed that the first term above tends uniformly in probability to $H(t, z)$. Using Doob's inequality, the expectation of the $\|\cdot\|$-norm of the second term is bounded above by

$$n w_n \int_0^1 E \sup_t \left| \int_0^t \frac{M^{(n)}(ds, x)}{(Y^{(n)}(s, x))^2} \right| dx \leqq 2 n w_n \|\alpha\| \left(\sup_{s, x} E\left(\frac{1}{Y^{(n)}(s, x)}\right)^3\right)^{1/2}$$
$$= O\left(\frac{1}{\sqrt{n w_n}}\right) \to 0.$$

by Condition (A3) and $n w_n \to \infty$. □

Proof of Proposition 3.3. Note that $\hat{H}$ can be considered as a piecewise linear approximation (in z for any fixed t) to the estimator

$$\tilde{H}(t, z) = n w_n^2 \sum_{r=1}^{[z d_n]} \int_0^t \frac{dX_r^{(n)}(s)}{(Y_r^{(n)}(s))^2},$$

where $X_r^{(n)}(s) = X^{(n)}(s, x_r)$. We first show that the L^1-distance between $\hat{h}$ and the estimator $\tilde{h}$, defined by replacing $\hat{H}$ by $\tilde{H}$ in $\hat{h}$, is negligible. Denote the locations of the jumps of the step function $\tilde{H}$ by (τ_j, x_r), and the corresponding jump sizes by Δ_{jr}. Then

$$\tilde{h}(t, z) = \frac{1}{b_n^2} \sum_{j, r} K\left(\frac{t-\tau_j}{b_n}\right) K\left(\frac{z-x_r}{b_n}\right) \Delta_{jr},$$
$$\hat{h}(t, z) = \frac{1}{b_n^2} \sum_{j, r} K\left(\frac{t-\tau_j}{b_n}\right) \frac{1}{w_n} \int_{\mathscr{I}_r} K\left(\frac{z-x}{b_n}\right) dx\, \Delta_{jr}.$$

Hence, using the Lipschitz condition on K,

$$\iint_{[0,1]^2} |\hat{h}(t, z) - \tilde{h}(t, z)|\, dt\, dz$$
$$\leqq \frac{1}{b_n^2} \sum_{j, r} \iint_{[0,1]^2} K\left(\frac{t-\tau_j}{b_n}\right) \left| K\left(\frac{z-x_r}{b_n}\right) - \frac{1}{w_n} \int_{\mathscr{I}_r} K\left(\frac{z-x}{b_n}\right) dx \right| dt\, dz\, \Delta_{jr}$$
$$\leqq C \frac{w_n}{b_n^2} \sum_{j, r} \Delta_{jr} = C \frac{w_n}{b_n^2} \tilde{H}(1, 1).$$

By the proofs of Lemmas 4 and 7 and Lemma 1 (c), $E\tilde{H}(1, 1) = O(1)$, so that, using $w_n = o(b_n^2)$, we obtain $E \iint_{[0,1]^2} |\hat{h}(t, z) - \tilde{h}(t, z)|\, dt\, dz \to 0$. It remains to prove the result for $\tilde{h}$. Write

(5.15) $$h - \tilde{h} = (h - h^0) + (h^0 - h^\dagger) + (h^\dagger - h^*) - R,$$

where

$$h^0(t,z)=\frac{1}{b_n^2}\iint_{[0,1]^2} K\left(\frac{t-s}{b_n}\right)K\left(\frac{z-x}{b_n}\right)h(s,x)\,ds\,dx,$$

$$h^\dagger(t,z)=\frac{1}{b_n^2 d_n}\sum_{r=1}^{d_n} K\left(\frac{z-x_r}{b_n}\right)\int_0^1 K\left(\frac{t-s}{b_n}\right)h(s,x_r)\,ds,$$

$$h^*(t,z)=\frac{1}{b_n^2 d_n}\sum_{r=1}^{d_n} K\left(\frac{z-x_r}{b_n}\right)\int_0^1 K\left(\frac{t-s}{b_n}\right)n w_n\frac{\alpha_r^{(n)}(s)}{(Y_r^{(n)}(s))^2}\,ds,$$

$$R(t,z)=\frac{n}{b_n^2 d_n^2}\sum_{r=1}^{d_n} K\left(\frac{z-x_r}{b_n}\right)\int_0^1 K\left(\frac{t-s}{b_n}\right)\frac{dM_r^{(n)}(s)}{(Y_r^{(n)}(s))^2},$$

and $\alpha_r^{(n)}(s)=\alpha_r^{(n)}(s,x_r)$. Now let us treat each term in (5.15) separately. First, since h is continuous, $\iint_{[0,1]^2}|h(t,z)-h^0(t,z)|\,dt\,dz\to 0$. Secondly, since h is continuous and K is Lipschitz,

$$\sup_{t,z}|h^0(t,z)-h^\dagger(t,z)|\leqq\frac{1}{b_n}\sup_{s,z}\left|\int_0^1 K\left(\frac{z-x}{b_n}\right)h(s,x)\,dx-\frac{1}{d_n}\sum_{r=1}^{d_n}K\left(\frac{z-x_r}{b_n}\right)h(s,x_r)\right|$$

$$\leqq\frac{1}{b_n}\sup_{s,z}\left\{\sum_{r=1}^{d_n}\int_{\mathscr{I}_r}\left|K\left(\frac{z-x}{b_n}\right)-K\left(\frac{z-x_r}{b_n}\right)\right|dx\,h(s,x_r)\right\}+o(1)$$

$$\leqq\frac{1}{b_n}O\left(\frac{1}{b_n d_n}\right)+o(1)\to 0,$$

by $w_n=o(b_n^2)$. Thirdly, using arguments similar to those in the proof of Lemma 4, the assumption that K has compact support and Lemma 1 (c),

$$\sup_{t,z}E|h^\dagger(t,z)-h^*(t,z)|\leqq\frac{1}{b_n^2 d_n}\sup_t\left[\int_0^1 K\left(\frac{t-s}{b_n}\right)ds\right]\sup_z\left[\sum_{r=1}^{d_n}K\left(\frac{z-x_r}{b_n}\right)\right]$$

$$\cdot\sup_{s,r}E\left|h(s,x_r)-n w_n\frac{\alpha_r^{(n)}(s)}{(Y_r^{(n)}(s))^2}\right|$$

$$=\frac{1}{b_n^2 d_n}O(b_n)\,O(b_n d_n)\,o(1)\to 0.$$

Finally, using the orthogonality of the martingales $M_r^{(n)}$, $r=1,\ldots,d_n$, and Lemma 1 (a)

$$\sup_{z,t}E|R(t,z)|^2=\frac{n^2}{b_n^4 d_n^4}\sup_{t,z}\left\{\sum_{r=1}^{d_n}K^2\left(\frac{z-x_r}{b_n}\right)\int_0^1 K^2\left(\frac{t-s}{b_n}\right)E\left[\frac{\alpha_r^{(n)}(s)}{(Y_r^{(n)}(s))^4}\right]ds\right\}$$

$$\leqq\frac{n^2}{b_n^4 d_n^4}O(b_n d_n)\,O(b_n)\,O\left(\frac{d_n}{n}\right)^3$$

$$=O\left(\frac{1}{n w_n}\right)\frac{w_n}{b_n^2}\to 0,$$

since $n w_n\to\infty$ and $w_n=o(b_n^2)$. □

References

Aalen, O.O.: Weak convergence of stochastic integrals related to counting processes. Z. Wahrscheinlichkeitstheor. Verw. Geb. **38**, 261–277 (1977) Correction: **48**, 347 (1979)

Aalen, O.O.: Nonparametric inference for a family of counting processes. Ann. Stat. **6**, 701–726 (1978)

Andersen, P.K., Gill, R.D.: Cox's regression model for counting processes: a large sample study. Ann. Stat. **10**, 1100–1120 (1982)

Bass, R.F.: Probability estimates for multiparameter Brownian processes. Ann. Probab. **16**, 251–264 (1988)

Beran, R.: Nonparametric regression with randomly censored survival data. Tech. Rep., Dept. of Stat., University of California, Berkeley 1981

Bickel, P.J., Wichura, M.J.: Convergence criteria for multiparameter stochastic processes and some applications. Ann. Stat. **42**, 1656–1670 (1971)

Billingsley, P.: Convergence of Probability Measures. New York: Wiley 1968

Billingsley, P.: Probability and measure. 2nd ed. New York: Wiley 1986

Cairoli, R.: Une inégalité pour martingales à indices multiples et ses applications. Séminaire de Probabilités IV. (Lect. Notes Math., vol. 124, pp. 1–27 Berlin Heidelberg New York: Springer 1970

Cox, D.R.: Regression models and life tables (with discussion). J. R. Stat. Soc. Ser B **34**, 187–220 (1972)

Dabrowska, D.M.: Nonparametric regression with censored survival time data. Scand. J. Stat. **14**, 181–197 (1987)

Dellacherie, C., Meyer, P.-A.: Probabilities and potential B. Amsterdam: North Holland 1982

Greenwood, P.E., Wefelmeyer, W.: Efficient estimation in a nonlinear counting process regression model. Preprint in Statistics no. 121, University of Cologne, 1989

Hall, W.J., Wellner, J.A.: Confidence bands for a survival curve from censored data. Biometrika **67**, 133–143 (1980)

Hastie, T.J., Tibshirani, R.J.: Generalized additive models (with discussion). Stat. Sci. **1**, 297–310 (1986)

Ito, K.: Multiple Wiener integral. J. Math. Soc. Japan **3**, 157–169 (1951)

McKeague, I.W.: A counting process approach to the regression analysis of grouped survival data. Stochastic Processes Appl. **28**, 221–239 (1988)

McKeague, I.W., Utikal, K.J.: Goodness-of-fit tests for additive hazards and proportional hazards models. Tech. Rep., Dept. of Stat. Florida State University, Tallahassee, 1988

McKeague, I.W., Utikal, K.J.: (1990a). Stochastic calculus as a tool in survival analysis: a review. Appl. Math. Comput. (to appear)

McKeague, I.W., Utikal, K.J.: (1990b). Inference for a nonlinear counting process regression model. Ann. Stat. (to appear)

Neuhaus, G.: On weak convergence of stochastic processes with multidimensional time parameter. Ann. Math. Stat. **42**, 1285–1295 (1971)

O'Sullivan, F.: Nonparametric estimation in the Cox proportional hazards model. Tech. Rep. Dept. of Stat. University of California, Berkeley, 1986a

O'Sullivan, F.: Relative risk estimation. Tech. Rep., Dept. of Stat. University of California, Berkeley (1986b)

Rebolledo, R.: Central limit theorems for local martingales. Z. Wahrscheinlichtstheor. Verw. Geb. **51**, 269–286 (1980)

Thomas, D.C.: Nonparametric estimation and tests of fit for dose response relations. Biometrics **39**, 263–268 (1983)

Tibshirani, R.J.: Local likelihood estimation. Tech. Rep. and unpublished Ph.D. dissertation, Dept. of Stat. Stanford University, 1984

Wong, E., Zakai, M.: Martingales and stochastic integrals for processes with a multidimensional parameter. Z. Wahrscheinlichkeitstheor. Verw. Geb. **51**, 109–122 (1974)

Ylvisaker, D.: A note on the absence of tangencies in Gaussian sample paths. Ann. Math. Stat. **39**, 261–262 (1968)