NONPARAMETRIC REGRESSION IN PROPORTIONAL HAZARDS MODELS

Toshio Honda*

Fan et al. (1997) considered two kinds of nonparametric estimators of the effects of the covariates in proportional hazards models. One of them has no parametric assumption on the baseline hazard function and is based on the integration of the estimated first order derivative of the regression function. We study the asymptotic properties of the estimator and consider another nonparametric estimator of the effects of the covariates in proportional hazards models. We show both of the estimators have very similar asymptotic properties. The latter is closely related to estimation in two-sample problems and is much easier to calculate.

Key words and phrases: Asymptotic distribution, censoring time, estimation of the first order derivative, failure time, integration, local partial likelihood, local polynomial fitting, nonparametric regression, proportional hazards models, two-sample problem.

1. Introduction

In this paper we consider nonparametric estimation of the effects of covariates in proportional hazards models under censoring. Suppose that we are interested in examining the relationship between a failure time $T$ and a 1-dimensional covariate $\{X(t)\}$ in a proportional hazards model and we have no idea as to the parametric form of the regression function with respect to $X(t)$. If a linear model is used for the effect of $X(t)$, the model is reduced to the Cox regression model in Cox (1972).

Suppose we have $n$ independent and identical observations of $(\delta, X(t), T \wedge C)$ on $[0, \tau]$, $(\delta_i, X_i(t), T_i \wedge C_i)$, $i = 1, \ldots, n$, where $C$ is a censoring time, $\delta = I(T \leq C)$, and $\tau < \infty$. We assume that the censoring time $C$ is conditionally independent of the failure time $T$ given $\{X(t)\}$.

Defining $N(t)$ and $Y(t)$ by $N(t) = I(T \leq t \wedge C)$ and $Y(t) = I(T \wedge C \geq t)$, we assume that there exists a filtration $\{\mathcal{F}_t \mid t \in [0, \tau]\}$ such that $\{N_i(t)\}$ is adapted and $\{X_i(t)\}$ and $\{Y_i(t)\}$ are predictable. Let the filtration satisfy the usual conditions as in Dabrowska (1997).

Then our model is specified by assuming that the intensity process of $\{N_i(t)\}$ on $[0, \tau]$ is

$$\Lambda_i(dt) = E\{N_i(dt) \mid \mathcal{F}_{t-}\} = Y_i(t) \exp(\psi(X_i(t))) \lambda_0(t) dt,$$

where $\psi(\cdot)$ is an unknown smooth function and $\lambda_0(\cdot)$ is an unknown bounded function. Suppose that there exists a standard point for the covariate and let
0 and $x$ be the standard point and a fixed value of the covariate, respectively. Then we assume that $\psi(\cdot)$ is twice continuously differentiable around 0 and $x$ in Section 2, and that $\psi(\cdot)$ is three times continuously differentiable on an open interval containing $[0, x]$ in Section 3.

We consider nonparametric estimation of $\psi(x) - \psi(0)$ because no parametric assumption is imposed on the baseline hazard function and $\psi(\cdot)$ is only identifiable up to a constant addition. We set $\psi(0) = 0$ for simplicity of presentation. Then $\lambda_0(\cdot)$ is the baseline hazard function. Since the censoring time $C$ is assumed to be conditionally independent of the failure time $T$, the intensity process of $\{I(T_i \leq t)\}$ is (1.1) with $Y_i(t)$ replaced by $I(T_i \geq t)$, and the effect of the covariate is still $\exp(\psi(X_i(t)))$. When we have (1.1), $\{M_i(t)\}$ defined by

$$M_i(t) = N_i(t) - \int_0^t Y_i(t) \exp(\psi(X_i(s))) \lambda_0(s) ds$$

is a martingale process with respect to $\{\mathcal{F}_t\}$, which is crucial to studying estimators in proportional hazards models. See Chapters 2 and 3 of Andersen et al. (1993) for the mathematical treatment of these subjects.

We deal with two nonparametric estimators of $\psi(x) = \psi(x) - \psi(0)$ in this paper. We call the estimator of Section 2 two-sample estimator and that of Section 3 of Fan et al. (1997) integration estimator since the former is related to estimation in an imaginary two-sample problem and the latter is based on the numerical integration of the estimated $\psi'(\cdot)$ over $[0, x]$. Both estimators are based on local partial likelihood and local polynomial fitting. We derive the asymptotic distributions in Sections 2 and 3.

The results of this paper imply that the rate of convergence of the integration estimator is better than that of the estimated $\psi'(\xi)$, $\xi \in [0, x]$, and worse than the optimal rate $n^{-3/7}$, and that both estimators are very similar in their asymptotic properties. Since the two-sample estimator is much easier to calculate and it requires the less restrictive smoothness assumption, the results are not favorable to the integration estimator. However, it is very important to present the results since several papers, such as Fan and Gijbels (2000) and Linton et al. (2003), refer to the integration estimator without the asymptotic properties. Note that Fan et al. (1997) gave a theoretical result on the estimator of $\psi'(\xi)$, not the integration estimator.

There are many papers on nonparametric estimation of hazard functions. As for proportional hazards models, for example, O’Sullivan (1993) studied smoothing splines, Fan et al. (1997) considered two kernel estimators, and Huang et al. (2000) considered nonparametric ANOVA models by using regression splines. One of the estimators in Fan et al. (1997) is for the cases of parametric baseline hazard functions. Nielsen et al. (1998) considered a similar problem. The other estimator of Fan et al. (1997) is the integration estimator in this paper and for the cases of nonparametric baseline hazard functions. Linton et al. (2003) considered additive models in proportional hazards models and proposed estimators of the component functions including the baseline hazard function. Honda (2003) considered a similar problem by applying the two-sample estimators.
Some authors considered nonparametric or semiparametric estimation of hazard functions without any proportionality assumptions on hazard functions. For example, Kooperberg et al. (1995) considered regression splines, Li and Doss (1995) studied kernel and nearest neighbor estimators, and Nielsen and Linton (1995) considered kernel estimators. Dabrowska (1997) studied a kind of partially linear models and the results are useful to our purpose.

When we have no proportionality assumption, the order of estimation deteriorates because of the effect of one additional covariate of time.

The two-sample estimator has some desirable properties. For example, the regression spline estimators, the smoothing spline estimators, and the integration estimators may have difficulty if covariates are sparsely observed over subsets of the interval in question. Suppose that the covariate is time independent and that we estimate \( \psi(x) - \psi(0) \). Then, if no covariate is observed between \( a \) and \( b \), where \( 0 < a < b < x \), the integration estimator cannot be used and both of the spline estimators may need some remedies. On the other hand, the two-sample estimator can be applied if we have observed covariates both around 0 and around \( x \). It also will help researchers check if the assumption of a linear function of the popular Cox regression is adequate.

Section 2 deals with the two-sample estimator and the asymptotic distribution is given in Theorem 1. The extension to the cases where \( X(t) \) is \( m \)-dimensional is straightforward. The integration estimator is studied in Section 3 and the asymptotic distribution is given in Theorem 2. In Section 3 we deal with the cases where the covariate is time independent. We give a remark on the extension to the cases of time dependent covariates at the end of Section 3.

2. Two-sample estimator

We apply the principle of the local partial likelihood to the nonparametric estimation of \( \psi(x) - \psi(0) \) as in Section 3 of Fan et al. (1997). However, we apply it in a more direct manner and construct another kind of estimator. Then we state the asymptotic distribution of the estimator in Theorem 1. Although the proof is essentially contained in that of Theorem 2 or in Honda (2003), we describe the outline for reference.

We follow Dabrowska (1997) and Fan et al. (1997) in examining the asymptotic properties of the estimator and present the asymptotic distribution of the two-sample estimator in Theorem 1 below.

Before defining the two-sample estimator, we describe assumptions and notations. Note that we assume that the setup around (1.1) holds throughout this paper. The following assumptions A1–2 are usual ones in the literature on nonparametric function estimation.

**Assumption A1.** The bounded and symmetric nonnegative kernel function \( K(\cdot) \) has the support contained in \([-1, 1]\). Besides \( C_{K_0} = 1 \), where \( C_{K_0} \) is defined in (2.1) below.

**Assumption A2.** The bandwidth \( h \) tends to 0 as \( n \to \infty \). In addition we
have $nh \to \infty$.

We define $C_{Kj}$, $C_{Kj}(\eta)$, $D_{Kj}$, and $K_h(\xi)$ by

\begin{align}
(2.1) & \quad C_{Kj} = \int u^j K(u) du, \quad C_{Kj}(\eta) = \int \exp(\eta u) u^j K(u) du, \\
& \quad D_{Kj} = \int u^j K^2(u) du, \quad K_h(\xi) = \frac{1}{h} K\left(\frac{\xi}{h}\right),
\end{align}

where $\eta \in \mathbb{R}$ and $\xi \in \mathbb{R}$. Note that $C_{K1} = C_{K3} = D_{K1} = 0$. This will be used in evaluating the asymptotic biases and variances.

Next we define $\tilde{X}(t)$, $\tilde{K}_h(X(t))$, $\tilde{N}(t)$, $\tilde{M}(t)$, and $\beta_j$, $j = 0, 1, 2, 3$. We define the localized covariate $\tilde{X}(t)$ by

\begin{align}
(2.2) & \quad \tilde{X}(t) = \begin{cases} 
(0, X(t)/h, 0)^T, & K_h(X(t)) > 0, \\
(1, 0, (X(t) - x)/h)^T, & K_h(X(t) - x) > 0, \\
(0, 0, 0)^T, & \text{otherwise},
\end{cases}
\end{align}

where $v^T$ means the transposed vector.

Let the covariate be time independent temporarily only in this paragraph for explanatory purposes. Then the two samples of the imaginary two-sample problem are one sample consisting of observations with $|X_i| \leq h$ and another sample consisting of observations with $|X_i - x| \leq h$. The first line in (2.2) is for one of the imaginary two samples $\{i \mid |X_i| \leq h\}$ and the second line is for the other sample $\{i \mid |X_i - x| \leq h\}$. The first element of the second line is the dummy variable in the two-sample problem.

When $K_h(X(t))K_h(X(t) - x) > 0$, $\tilde{X}(t)$ is defined by the first line in (2.2) if $X(t)$ is closer to 0 and by the second line if $X(t)$ is closer to $x$. Hereafter let $n$ be large enough to have $|x - 0| > 2h$.

Besides, we put

\begin{align}
(2.3) & \quad \tilde{K}_h(X(t)) = K_h(X(t)) + K_h(X(t) - x), \\
(2.4) & \quad \tilde{N}(t) = N(t)\tilde{K}_h(X(t)), \\
(2.5) & \quad \tilde{M}(t) = \tilde{N}(t) - \int_0^t Y(s)\tilde{K}_h(X(s)) \exp(\psi(X(s)))\lambda_0(s) ds.
\end{align}

As is $\{M_i(t)\}$, $\{\tilde{M}_i(t)\}$ is a martingale process with respect to $\{\mathcal{F}_t\}$.

In order to approximate $\psi(X(t))$ by using the Taylor expansion, we define $\beta_j$, $j = 0, 1, 2, 3$ by

\begin{align}
(2.6) & \quad \beta_0 = (\beta_1, \beta_2, \beta_3)^T, \quad \beta_1 = \psi(x), \quad \beta_2 = h\psi'(0), \quad \beta_3 = h\psi'(x).
\end{align}

We approximate $\psi(X(t))$ by $\beta_0^\top \tilde{X}(t)$ and apply the principle of local partial likelihood. Then we estimate $\beta_0$ by maximizing the local partial likelihood. Note that $\beta_2$ and $\beta_3$ tend to 0 as $n \to \infty$. 

We define \( S^{(j)}(t, \eta) \), \( j = 0, 1, 2 \), as in Dabrowska (1997).

\[
S^{(0)}(t, \eta) = \sum_{i=1}^{n} Y_i(t) \exp \left( \eta^T \tilde{X}_i(t) \right) \tilde{K}_h(X_i(t)),
\]

\[
S^{(1)}(t, \eta) = \sum_{i=1}^{n} Y_i(t) \tilde{X}_i(t) \exp \left( \eta^T \tilde{X}_i(t) \right) \tilde{K}_h(X_i(t)),
\]

\[
S^{(2)}(t, \eta) = \sum_{i=1}^{n} Y_i(t)(\tilde{X}_i(t))^\otimes 2 \exp \left( \eta^T \tilde{X}_i(t) \right) \tilde{K}_h(X_i(t)),
\]

where \( \eta \) is a 3 \( \times \) 1 vector and \( v^{\otimes 2} = vv^T \) for an \( m \times 1 \) vector \( v \).

We use observations with \( |X_i(t)| \leq h \) or \( |X_i(t) - x| \leq h \) and follow the principle of partial likelihood by employing the local linear approximation of \( \psi(\cdot) \). Then we get the local partial likelihood, which is written as

\[
L(\beta) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left( \beta^T \tilde{X}_i(t) - \log S^{(0)}(t, \beta) \right) \tilde{N}_i(dt),
\]

where \( \beta \) is a 3 \( \times \) 1 vector. See Fan et al. (1997) for the derivation of (2.10). We estimate \( \beta_0 \) by maximizing (2.10), i.e., we choose as the estimate of \( \beta_0 \) a 3 \( \times \) 1 vector \( \hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^T \) such that

\[
L(\hat{\beta}) = \max L(\beta).
\]

We estimate \( \psi(x) \) by \( \hat{\beta}_1 \).

Since

\[
\frac{\partial^2 L}{\partial \eta \partial \eta^T}(\eta) = -\frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left( \frac{S^{(2)}(t, \eta)}{S^{(0)}(t, \eta)} - \frac{(S^{(1)}(t, \eta))^{\otimes 2}}{(S^{(0)}(t, \eta))^2} \right) \tilde{N}_i(dt),
\]

\( L(\eta) \) will be strictly concave unless \( \sum_{i=1}^{n} \tilde{N}_i(\tau) \) is very small. Then the estimate \( \hat{\beta} \) will be uniquely defined.

The consistency of \( \hat{\beta} \) can be proved in the same way as in Section 6.3 of Fan et al. (1997).

When we derive the asymptotic distribution of \( \hat{\beta} - \beta_0 \), it is crucial that \( \hat{\beta} - \beta_0 \) is written as

\[
\hat{\beta} - \beta_0 = \left(-\frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta^*)\right)^{-1} U(\beta_0),
\]

where \( \beta^* = \beta_0 + \theta(\hat{\beta} - \beta_0) \), \( 0 < \theta < 1 \), and \( U(\beta) = (\partial L/\partial \beta)(\beta) \), which corresponds to the efficient score function of the Cox regression model.

Before stating Theorem 1, we give more notations and assumptions. Assumptions A3–4 are standard for nonparametric estimation in proportional hazards models and necessary for Assumption A5. Assumption A5 is the same as I.6 in
Dabrowska (1997). Comments on the assumption are given in Dabrowska (1997). It is also easy to see that Assumptions A3–4 and the Hölder continuity of the covariate, \(|X(t) - X(s)| < L|t - s|^\gamma\), for some constants \(L\) and \(\gamma\), are enough for Assumption A5. The monotonicity of \(\{Y(t)\}\) is crucial. The Hölder continuity can be relaxed by strengthening the other assumptions. See Remark 1 at the end of Section 3.

**Assumption A3.** Denote the density function of \(X(t)\) by \(f(\cdot, t)\). Then \(f(0, t) > \epsilon_1\) and \(f(x, t) > \epsilon_1\) on \([0, \tau]\) for a positive \(\epsilon_1\). In addition \(f(\xi, t)\) is continuous on \([-\epsilon_2, \epsilon_2] \times [0, \tau]\) and \([x - \epsilon_2, x + \epsilon_2] \times [0, \tau]\) for a positive \(\epsilon_2\).

**Assumption A4.** Denote \(E(Y(t) \mid X(t) = \xi)\) by \(g(\xi, t)\). Then \(g(\xi, t)\) is continuous on \([-\epsilon_3, \epsilon_3] \times [0, \tau]\) and \([x - \epsilon_3, x + \epsilon_3] \times [0, \tau]\) for a positive \(\epsilon_3\).

Hereafter, when we refer to the uniformity of convergence in \(\beta\) or \(\eta\), we mean the uniformity in \(\beta\) or \(\eta\) on \{\(|\beta| < M\)\} or \{\(|\eta| < M\)\} for a sufficiently large \(M\). The uniformity in \(t\) means that on \([0, \tau]\).

**Assumption A5.** We have the following uniform convergence in probability of \(S^{(i)}\) in \(t\) and \(\eta\).

\[
\begin{align*}
\frac{1}{n} S^{(0)}(t, \eta) & \to f(0, t)g(0, t)C_{K0}(\eta_2) + f(x, t)g(x, t)e^{\eta}C_{K0}(\eta_3), \\
\frac{1}{n} S^{(1)}(t, \eta) & \to f(0, t)g(0, t)(0, C_{K1}(\eta_2), 0)^T \\
& \quad + f(x, t)g(x, t)e^{\eta}(C_{K0}(\eta_3), 0, C_{K1}(\eta_3))^T, \\
\frac{1}{n} S^{(2)}(t, \eta) & \to f(0, t)g(0, t) \begin{pmatrix} 0 & 0 & 0 \\ 0 & C_{K2}(\eta_2) & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
& \quad + f(x, t)g(x, t)e^{\eta} \begin{pmatrix} C_{K0}(\eta_3) & 0 & C_{K1}(\eta_3) \\ 0 & 0 & 0 \\ C_{K1}(\eta_3) & 0 & C_{K2}(\eta_3) \end{pmatrix},
\end{align*}
\]

where \(\eta = (\eta_1, \eta_2, \eta_3)^T\).

We define \(V(\xi), W(x, t), \) and \(\tilde{V}\) appearing in Theorem 1.

\[
\begin{align*}
V(\xi) & = \int_0^\tau f(\xi, t)g(\xi, t)\exp(\psi(\xi))\lambda_0(t)dt, \\
W(x, t) & = f(0, t)g(0, t) + f(x, t)g(x, t)e^{\beta_1}, \\
\tilde{V} & = \int_0^\tau (W(x, t))^{-1}f(0, t)g(0, t)f(x, t)g(x, t)e^{\beta_1}\lambda_0(t)dt.
\end{align*}
\]

**Theorem 1.** In addition to the setup around (1.1), we assume Assumptions A1–5. Then as \(n \to \infty\), \((nh)^{1/2}(\beta - \beta_0 - \text{Bias})\) converges in law to \(N(0, \Sigma(x))\),
where
\[
\Sigma(x) = \begin{pmatrix}
D_{K0} \bar{V}^{-1} & 0 & 0 \\
0 & D_{K2}(C_{K2}^2 V(0))^{-1} & 0 \\
0 & 0 & D_{K2}(C_{K2}^2 V(x))^{-1}
\end{pmatrix},
\]
\[
Bias = \left(\frac{h^2 C_{K2}}{2} \left(\psi''(x) - \psi''(0)\right) + o(h^2), o(h^2), o(h^2)\right)^T.
\]

Theorem 1 shows that the asymptotically optimal bandwidth for \(\hat{\psi}(x)\) is \(cn^{-1/5}\), where \(c\) depends on \(\bar{V}, \psi''(0), \) and \(\psi''(x)\). For example, the asymptotically optimal bandwidth in the sense of the asymptotic MSE is given by
\[
h = n^{-1/5} \left(\frac{D_{K0}}{V C_{K2}^2 (\psi''(x) - \psi''(0))^2}\right)^{1/5}.
\]

Then the estimator achieves the optimal convergence rate in the sense of the asymptotic MSE, \(n^{-4/5}\), since we assume that \(\psi(\cdot)\) is only twice continuously differentiable in this section.

Local quadratic fitting is applied to estimate \(\psi''(0)\) and \(\psi''(x)\) as in Fan et al. (1997). We can also estimate \(\bar{V}\) by
\[
\int_0^\tau \left\{ \left(\frac{1}{n} \sum_{i=1}^n Y_i(t) K_h(X_i(t))\right)^{-1} + \left(\frac{1}{n} \sum_{i=1}^n Y_i(t) K_h(X_i(t) - x)e^{\hat{\psi}(x)}\right)^{-1} \right\}^{-1} \frac{\sum_{i=1}^n N_i(dt)}{\sum_{i=1}^n Y_i(t) \exp(\hat{\psi}(X_i(t)))}.
\]

We need to use a rule of thumb to choose bandwidths for \(\hat{\psi}(\cdot)\) in (2.14).

**Proof of Theorem 1.** At first we decompose \(U(\beta_0)\) in (2.12) into
\[
U(\beta_0) = U_1(\beta_0) + U_2(\beta_0)
\]
\[
= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\bar{X}_i(t) - \frac{S^{(1)}(t, \beta_0)}{S^{(0)}(t, \beta_0)}\right) \tilde{M}_i(dt)
\]
\[
+ \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(t) \left(\bar{X}_i(t) - \frac{S^{(1)}(t, \beta_0)}{S^{(0)}(t, \beta_0)}\right) \tilde{K}_h(X_i(t)) e^{\hat{\psi}(X_i(t))} \lambda_0(t) dt.
\]
We begin with $U_2(\beta_0)$, which is related to the bias term in Theorem 1. As in Fan et al. (1997), we have

$$
U_2(\beta_0) = \frac{1}{n} \sum_{i=1}^{n} \int_0^T Y_i(t) \left( \tilde{X}_i(t) - \frac{S^{(1)}(t, \beta_0)}{S^{(0)}(t, \beta_0)} \right) \times \tilde{K}_h(X_i(t)) \left( e^{\psi(X_i(t))} - e^{\beta_0^T \tilde{X}_i(t)} \right) \lambda_0(t) dt.
$$

By the Taylor expansion, we have

$$
\tilde{K}_h(X_i(t)) \left( e^{\psi(X_i(t))} - e^{\beta_0^T \tilde{X}_i(t)} \right)
$$

$$
= \frac{h^2}{2} \left\{ K_h(X_i(t)) \psi''(0) \left( \frac{X_i(t)}{h} \right)^2 + K_h(X_i(t) - x) \psi''(x) e^{\beta_1} \left( \frac{X_i(t) - x}{h} \right)^2 \right\} + o(h^2).
$$

From (2.16), (2.17), and Assumptions A3–5 with $\eta = \beta_0$, we get

$$
\frac{U_2(\beta_0)}{h^2} \rightarrow \left( \frac{1}{2} (\psi''(x) - \psi''(0)) C_{K^2 V}, 0, 0 \right)^T
$$

in probability.

Next we deal with $U_1(\beta_0)$, which is related to the convergence in law to the normal distribution in Theorem 1. Since the proof of the asymptotic normality of $(nh)^{1/2} U_1(\beta_0)$ is almost the same as in Fan et al. (1997) and Dabrowska (1997), we omit the details. We just note that the predictable variation process of

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t Y_i(s) \left( \tilde{X}_i(s) - \frac{S^{(1)}(s, \beta_0)}{S^{(0)}(s, \beta_0)} \right) \tilde{M}_i(ds)
$$

is given by

$$
\frac{1}{n} \sum_{i=1}^{n} \int_0^t Y_i(s) \left( \tilde{X}_i(s) - \frac{S^{(1)}(s, \beta_0)}{S^{(0)}(s, \beta_0)} \right)^{\otimes 2} \tilde{K}_h(X_i(s)) e^{\psi(X_i(s))} \lambda_0(s) ds.
$$

Hence we have

$$
(nh)^{1/2} U_1(\beta_0) \rightarrow N(0, \Omega(x))
$$

in law, where $\Omega(x)$ is defined by

$$
\Omega(x) = \begin{pmatrix}
D_{K_0} \tilde{V} & 0 & 0 \\
0 & D_{K^2 V(0)} & 0 \\
0 & 0 & D_{K^2 V(x)}
\end{pmatrix}.
$$
Finally we consider

$$\frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta^*) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \frac{S^{(2)}(t, \beta^*)}{S^{(0)}(t, \beta^*)} \tilde{N}_i(dt)$$

$$- \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \frac{(S^{(1)}(t, \beta^*))^2}{(S^{(0)}(t, \beta^*))^2} \tilde{N}_i(dt).$$

Since $\beta^* - \beta_0 \to 0$ in probability, Assumptions A3–5 imply that the first term of the right hand side of (2.20) converges in probability to

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & C_{K2} & 0 \\ 0 & 0 & 0 \end{pmatrix} V(0) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C_{K2} \end{pmatrix} V(x).$$

The second term converges in probability to

$$- \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \int_0^\tau \frac{(f(x, t)g(x, t)e^{\beta_1})^2}{W(x, t)} \lambda_0(t) dt.$$  

Combining (2.21) and (2.22), we have

$$\frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta^*) \to \begin{pmatrix} \tilde{V} & 0 & 0 \\ 0 & C_{K2}V(0) & 0 \\ 0 & 0 & C_{K2}V(x) \end{pmatrix}$$

in probability.

The proof is complete from (2.12), (2.18), (2.19), and (2.23).

3. Integration estimator

We deal with the estimator of Section 3 of Fan et al. (1997), which we call the integration estimator. We present the asymptotic distribution of the integration estimator in Theorem 2 below and show that the optimal bandwidth is $c_n^{-1/5}$.

Note that we use a common $h$ for all the points between 0 and $x$ and that we consider the ideal estimate in (3.5) below, not the numerically integrated one.

We assume that the covariate is time independent for simplicity and we denote it by $X$ in this section. See Remark 1 at the end of this section for the cases of time dependent covariates.

(3.22) below implies that the convergence in law to the normal distribution in Theorem 2 almost depends only on the observations whose covariates are around 0 or $x$. In that sense the integration estimator can be thought of as a kind of two-sample estimator although it looks quite different and is much more complicated.

We introduce several assumptions. The setup around (1.1) in Section 1 holds in this section, too.
ASSUMPTION A6. In addition to Assumption A1, we assume that $K(\cdot)$ is Lipschitz continuous.

Assumption A6 is used to establish the uniform convergence in Lemma 1 and so on. When $K(\xi) = I(-1/2 < \xi < 1/2)$, we can obtain the same result as in Theorem 2 with more elaborate arguments on the uniformity of convergence.

ASSUMPTION A7. In addition to Assumption A2, $\log n/(nh^3) = o(1)$. This implies that $h^{-1}\{(nh)^{-1} \log n\}^{1/2} \to 0$.

ASSUMPTION A8. The density function $f(\cdot)$ of $X$ is twice continuously differentiable and positive on $(-\epsilon_1, x + \epsilon_1)$ for a positive $\epsilon_1$. Note that $X$ is time independent in this section.

ASSUMPTION A9. Denote $E(Y(t) | X = \xi)$ by $g(\xi, t)$. For a positive $\epsilon_2$, $g(\xi, t) > \epsilon_2$ on $[0, x] \times [0, \tau]$. There exists an open set of $R^2$ containing $[0, x] \times [0, \tau]$ on which $g(\xi, t)$ is twice continuously differentiable with respect to $\xi$ and continuously differentiable with respect to $t$. Besides all the derivatives are uniformly bounded on the open set.

Assumptions A8–9 are more restrictive than Assumptions A3–4 in Section 2.

To use the notations in Section 2 such as $L(\beta)$, $U(\beta_0)$, $U_j(\beta_0)$, and $S^{(j)}$, we modify the definitions of $\tilde{K}_h(X(t)) = \tilde{K}_h(X)$, $\tilde{X}(t) = \tilde{X}$, $\beta_0$, $N$, and $M$. See (2.2)–(2.10), and (2.12) in Section 2.

(3.1) $\tilde{K}_h(X) = K_h(X - \xi)$, $\tilde{M}(dt) = \tilde{K}_h(X)M(dt)$, $\tilde{N}(dt) = \tilde{K}_h(X)N(dt)$,

(3.2) $\beta_0 = (\beta_1, \beta_2)^T$, $\beta_1 = h\psi'(\xi)$, $\beta_2 = \frac{h^2}{2}\psi''(\xi)$,

(3.3) $\tilde{X} = \begin{cases} ((X - \xi)/h, (X - \xi)^2/h^2)^T, & K_h(X - \xi) > 0 \\ (0, 0)^T, & \text{otherwise} \end{cases}$.

Note that all the above definitions depend on $\xi \in [0, x]$ and that we consider local quadratic fitting. As for (3.3), even if we include a constant term in it, the constant term does not appear in the local partial likelihood.

We estimate $\beta_0$ as in Section 2 by maximizing the local partial likelihood $L(\beta)$ which is constructed from observations with $|X_i - \xi| \leq h$ by combining the local quadratic approximation and the principle of partial likelihood. We denote the estimate of $\beta_0$ by $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$, which maximizes the local partial likelihood. The local partial likelihood function will be strictly concave unless the sample size is very small.

Then we have for $\xi \in [0, \tau]$,

(3.4) $\hat{\beta} - \beta_0 = \left(-\frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta^*)\right)^{-1}(U_1(\beta_0) + U_2(\beta_0))$. 
where
\[
U(\beta_0) = U_1(\beta_0) + U_2(\beta_0) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \left( \tilde{X}_i - \frac{S^{(1)}(t, \beta_0)}{S^{(0)}(t, \beta_0)} \right) \tilde{M}_i(dt) + \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} Y_i(t) \left( \tilde{X}_i - \frac{S^{(1)}(t, \beta_0)}{S^{(0)}(t, \beta_0)} \right) \tilde{K}_h(X_i) \exp(\psi(X_i)) \lambda_0(t) dt.
\]

See also (2.12) in Section 2.

We estimate \(\psi(x) = \psi(x) - \psi(0)\) by numerically integrating \(\hat{\psi}'(\xi) = h^{-1}\hat{h}\psi'(\xi)\) over \([0, x]\), where \(\hat{h}\psi'(\xi)\) is \(\hat{\beta}_1\) for \(\xi\). In this paper we deal with the ideal \(\hat{\psi}(x) - \psi(x)\) and it is defined by
\[
(3.5) \quad \hat{\psi}(x) - \psi(x) = \frac{1}{h} \int_{0}^{x} \left( \hat{h}\psi'(\xi) - h\psi'(\xi) \right) d\xi.
\]

The asymptotic distribution of (3.5) is given in Theorem 2.

**Theorem 2.** In addition to the setup around (1.1), we assume Assumptions A6–9. Then as \(n \to \infty\), \((nh)^{1/2}(\hat{\psi}(x) - \psi(x) - \text{Bias})\) converges in law to \(N(0, \sigma^2(x))\), where
\[
\sigma^2(x) = \left( \frac{1}{V(0)} + \frac{1}{V(x)} \right) C_{K2}^{-2} \int_{-1}^{1} \left( \int_{z}^{1} uK(u)du \right)^2 dz,
\]
\[
\text{Bias} = \frac{h^2}{6C_{K2}} (\psi''(x) - \psi''(0)) + o(h^2),
\]
where \(V(\xi)\) is defined in (2.13) with \(f(\xi, t) = f(\xi)\).

Note again that we use a common \(h\) for every \(\xi\). The analysis will be extremely difficult if the bandwidths depend on \(\xi\). Theorem 2 is very similar to Theorem 1 in the convergence in law and the asymptotic bias. Theorem 2 shows that when a common \(h\) is used for every \(\xi\), the asymptotically optimal bandwidth is \(h = cn^{-1/5}\), where \(c\) depends on \(V(0), V(x), \psi''(0)\) and \(\psi''(x)\). For example, the asymptotically optimal bandwidth in the sense of the asymptotic MSE is given by
\[
h = n^{-1/5} \left( \frac{9\sigma^2(x)C_{K2}^2}{C_{K4}^2(\psi''(x) - \psi''(0))^2} \right)^{1/5}.
\]

Then the rate of convergence in the sense of the asymptotic MSE is \(n^{-4/5}\). This is better than that of \(\hat{\psi}'(\xi) = h^{-1}\hat{h}\psi'(\xi)\) and worse than the optimal rate \(n^{-6/7}\). See Theorem 4 in Fan et al. (1997) for the asymptotic properties of \(\hat{\psi}'(\xi)\). The evaluation of \(U_1(\beta_0)\) in the proof of the theorem implies that the effects of small values of \(f(\xi)\) may be serious.
Before we prove the theorem, we describe a lemma which is similar to Assumption A5 in Section 2. We should modify the definition of \(C_{K_j}(\eta)\) in (2.1) as

\[
C_{K_j}(\eta) = \int u^j \exp(\eta_1 u + \eta_2 u^2) K(u) du,
\]

where \(\eta = (\eta_1, \eta_2)^T\). Note that \(C_{K_j}(0) = C_{K_j}\) as in Section 2.

**Lemma 1.** Under the assumptions of Theorem 2, we have the following uniform convergence in probability of \(S^{(j)}\) in \(\eta, \xi,\) and \(t\).

\[
\frac{1}{n} S^{(0)}(t, \eta) = \frac{f(\xi)g(\xi, t)C_{K0}(\eta) + h(f'(\xi)g(\xi, t) + f(\xi)g_\xi(\xi, t))C_{K1}(\eta)}{n} + O_p((nh)^{-1} \log n)^{1/2} + O_p(h^2),
\]

\[
\frac{1}{n} S^{(1)}(t, \eta) = \frac{f(\xi)g(\xi, t)(C_{K1}(\eta), C_{K2}(\eta))^T + h(f'(\xi)g(\xi, t) + f(\xi)g_\xi(\xi, t))}{n} \times (C_{K2}(\eta), C_{K3}(\eta))^T + O_p((nh)^{-1} \log n)^{1/2} + O_p(h^2),
\]

\[
\frac{1}{n} S^{(2)}(t, \eta) = \frac{f(\xi)g(\xi, t)}{n} \left( C_{K2}(\eta) \ C_{K3}(\eta) \ C_{K4}(\eta) \right) + h(f'(\xi)g(\xi, t) + f(\xi)g_\xi(\xi, t)) \times \left( C_{K3}(\eta) \ C_{K4}(\eta) \ C_{K5}(\eta) \right) + O_p((nh)^{-1} \log n)^{1/2} + O_p(h^2),
\]

where \(g_\xi(\xi, t) = (\partial g/\partial \xi)(\xi, t)\).

The lemma is proved by following the arguments in Masry (1996) and using the monotonicity of \(\{Y(t)\}\) and Assumptions A6 and A8–9. The proof is omitted since the arguments are standard in the literature on nonparametric regression and the outline is given in Honda (2003).

**Proof of Theorem 2.** By establishing the uniform convergence of the partial likelihood \(L(\beta) - n^{-1} \log n \sum N_i(\tau)\) in \(\beta\) and \(\xi\), the uniform consistency of \(\hat{\beta}\) is proved and we have

\[
\hat{\beta} - \beta_0 = o_p(1),
\]

uniformly in \(\xi\). The proof of (3.6) is easier than the other part of that of Theorem 2 and omitted.

As in Section 2, the convergence in law comes from \(U_1(\beta_0)\) and the bias comes from \(U_2(\beta_0)\). The proof consists of 3 steps, the evaluation of \((\partial^2 L/\partial \beta \partial \beta^T)(\beta^*)_\), the evaluation of \(U_1(\beta_0)\), and the convergence in law to the normal distribution. We omit the details on the bias part since it is much easier.

Explicit expressions of \(\Omega_j\) \((j \geq 1)\), \(R_j\), and \(W_j\) \((j \geq 1)\) appearing later in the proof are not necessary since their effects are negligible in the asymptotic distribution. Only properties such as continuous differentiability are relevant in the proof. Thus we do not give the explicit expressions.
Step 1) At first we evaluate \( (\partial^2 L/\partial \beta \partial \beta^T)(\beta^*) \) in (3.4). It is given by

\[
- \frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta^*) = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left( \frac{S^{(2)}(t, \beta^*)}{S^{(0)}(t, \beta^*)} - \frac{(S^{(1)}(t, \beta^*))^{\otimes 2}}{(S^{(0)}(t, \beta^*))^2} \right) \tilde{N}_i(dt).
\]

From Lemma 1, we have uniformly in \( \beta \) and \( \xi \),

\[
(3.7) \quad - \frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta) = \Omega_0(\beta) \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \tilde{N}_i(dt) + h \sum_{j=1}^{5} \Omega_j(\beta) \times \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau W_j(\xi, t) \tilde{N}_i(dt) + O_p\left(\{(nh)^{-1} \log n\}^{1/2}\right) + O_p(h^2),
\]

where \( \Omega_0(\beta) \) is defined by

\[
\frac{1}{C_{K0}(\beta)} \begin{pmatrix} C_{K2}(\beta) & C_{K3}(\beta) \\ C_{K3}(\beta) & C_{K4}(\beta) \end{pmatrix} - \frac{1}{C_{K0}^2(\beta)} \begin{pmatrix} C_{K1}^2(\beta) & C_{K1}(\beta)C_{K2}(\beta) \\ C_{K1}(\beta)C_{K2}(\beta) & C_{K2}^2(\beta) \end{pmatrix},
\]

\( \Omega_j(\beta), 1 \leq j \leq 5, \) are 2 \times 2-dimensional deterministic functions and continuously differentiable, \( W_j(\xi, t), 1 \leq j \leq 5, \) are scalar deterministic functions and continuously differentiable with respect to \( \xi \), and \( \Omega_j, W_j, \) and all the derivatives are uniformly bounded.

Applying the arguments in Masry (1996) to (3.7), we have uniformly in \( \beta \) and \( \xi \),

\[
(3.8) \quad - \frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta) = \Omega_0(\beta)V(\xi) + h \sum_{j=1}^{5} \Omega_j(\beta) \int_0^\tau f(\xi) g(\xi, t) W_j(\xi, t) e^{\psi(\xi)} \lambda_0(t) dt + O_p\left(\{(nh)^{-1} \log n\}^{1/2}\right) + O_p\left(h^2\right).
\]

If we establish that uniformly in \( \xi \),

\[
(3.9) \quad |U_1(\beta_0)| + |U_2(\beta_0)| = O_p\left(\{(nh)^{-1} \log n\}^{1/2}\right) + O_p\left(h^2\right),
\]

(3.4), (3.8), and (3.9) imply that uniformly in \( \xi \),

\[
(3.10) \quad \beta^* - \beta_0 = O_p\left(\{(nh)^{-1} \log n\}^{1/2}\right) + O_p\left(h^2\right).
\]

See also (2.12). We verify (3.9) later in Step 2.

From (3.8) and (3.10),

\[
(3.11) \quad - \frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta^*) = \Omega_0(\beta_0)V(\xi) + h \sum_{j=1}^{5} \Omega_j(\beta_0) \int_0^\tau f(\xi) g(\xi, t) W_j(\xi, t) e^{\psi(\xi)} \lambda_0(t) dt + O_p\left(\{(nh)^{-1} \log n\}^{1/2}\right) + O_p\left(h^2\right) = \Omega_0(0)V(\xi) + h\Omega_6(\xi) + O_p\left(\{(nh)^{-1} \log n\}^{1/2}\right) + O_p\left(h^2\right),
\]
uniformly in \( \xi \), where \( \Omega_6(\xi) \) is a \( 2 \times 2 \)-dimensional deterministic function and continuously differentiable. We applied the Taylor expansion at \( \beta = (0, 0)^T \) and then used (3.2) to get the last expression in (3.11).

(3.11) implies that uniformly in \( \xi \),

\[
\left( -\frac{\partial^2 L}{\partial \beta \partial \beta^T}(\beta^*) \right)^{-1} = V^{-1}(\xi)\Omega_0^{-1}(0) + h\Omega_7(\xi) + O_p\left((nh)^{-1} \log n \right)^{1/2} + O_p(h^2),
\]

where \( \Omega_7(\xi) \) is a \( 2 \times 2 \)-dimensional deterministic function and continuously differentiable and

\[
\Omega_0^{-1}(0) = \begin{pmatrix} C_{K2}^{-1} & 0 \\ 0 & (C_{K4} - C_{K2}^2)^{-1} \end{pmatrix}.
\]

The third and fourth terms of the RHS of (3.12) are negligible by (3.9) and Assumption A7.

**Step 2)** We mainly consider \( U_1(\beta_0) \), which is written as

\[
U_1(\beta_0) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \tilde{X}_i \tilde{M}_i(dt) - \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \frac{S^{(1)}(t, \beta_0)}{S^{(0)}(t, \beta_0)} \tilde{M}_i(dt).
\]

We establish (3.9) and prove that the second term of the RHS of (3.14) is negligible in Theorem 2.

Noting that the summand of the first term of the RHS of (3.14) is \( \tilde{X}_i K_h(X_i - \xi) M_i(\tau) \) and following the arguments in Masry (1996), we have uniformly in \( \xi \),

\[
\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \tilde{X}_i \tilde{M}_i(dt) = O_p\left((nh)^{-1} \log n \right)^{1/2}.
\]

In addition we can prove that \( U_2(\beta_0) = O_p(h^2) \) uniformly in \( \xi \) as in Masry (1996). See also Section 6.3 of Fan et al. (1997).

If we show that the second term of the RHS of (3.14) is \( O_p((nh)^{-1} \log n)^{1/2}) + O_p(h^2) \), (3.9) is established.

By (3.2) and Lemma 1, the second term of the RHS of (3.14) is rewritten as

\[
\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \frac{S^{(1)}(t, \beta_0)}{S^{(0)}(t, \beta_0)} \tilde{M}_i(dt) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \tilde{M}_i(dt) + \frac{C_{K2}}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \tilde{M}_i(dt) + \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \Omega_8(\xi, t) \tilde{M}_i(dt) + \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} R(\xi, t) \tilde{M}_i(dt),
\]
where $\Omega_8(\xi, t)$ is a 2-dimensional deterministic function and continuously differentiable with respect to $\xi$, $\Omega_8(\xi, t)$ and its derivatives are uniformly bounded, and $R(\xi, t)$ is a 2-dimensional predictable process satisfying

\[ R(\xi, t) = O_p \left( \{(nh)^{-1} \log n\}^{1/2} \right) + O_p \left( h^2 \right), \]

uniformly in $t$ and $\xi$.

Since $\Omega_8(\xi, t)$ is smooth in $\xi$ and $R(\xi, t)$ satisfies (3.17), we get

\[ \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{T} \frac{S^{(1)}(t, \beta_0)}{S^{(0)}(t, \beta_0)} \tilde{M}_i(dt) = O_p \left( \{(nh)^{-1} \log n\}^{1/2} \right) + O_p \left( h^2 \right), \]

uniformly in $\xi$. Here we used the fact that

\[ \int_{0}^{T} \vert \tilde{M}_i \vert(dt) = O_p(1), \]

uniformly in $\xi$. Hence (3.9) is established.

Next we show that the effect of (3.16), the second term of the RHS of (3.14), is negligible in the asymptotic distribution in Theorem 2. Then we need the following facts (3.18) and (3.19) to evaluate it.

If $G(\xi, t)$ is a bounded deterministic function, we have

\[ \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{T} \left( \int_{0}^{x} G(\xi, t)K_h(X_i - \xi)d\xi \right) M_i(dt) = O_p \left( n^{-1/2} \right). \]

(3.18) is due to the evaluation of the second moment.

If $H(\xi, t) = O_p \left( \{(nh)^{-1} \log n\}^{1/2} \right) + O_p(h^2)$ uniformly in $t$ and $\xi$ and $H(\xi, t)$ is predictable, we have

\[ \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{T} \left( \int_{0}^{x} H(\xi, t)K_h(X_i - \xi)d\xi \right) M_i(dt) = O_p \left( \left\{ (n^2h^3)^{-1} \log n \right\}^{1/2} \right) + O_p \left( hn^{-1/2} \right). \]

(3.19) is due to Lenglart's inequality.

We multiply (3.16) by (3.12) from the left. Then divide the product by $h$ and integrate it with respect to $\xi$.

The effect of the first term of the RHS of (3.16) is $O_p(n^{-1/2})$ since $\Omega_0^{-1}(0)$ in (3.13) is diagonal and (3.18) applies to the product of the first term of the RHS of (3.16) and the second term of (3.12).

The effect of the second term of the RHS of (3.16) is also $O_p(n^{-1/2})$ from (3.18). The third term is negligible from (3.19) since $R(\xi, t)$ satisfies the conditions for (3.19). Thus (3.16) does not affect the asymptotic distribution of $\hat{\psi}(x) - \psi(x)$. 

**Step 3** We verify the convergence in law in Theorem 2. Step 2 implies that we should deal with only the first term in the RHS of (3.14).

We multiply it by (3.12) from the left. Then divide the product by $h$ and integrate it with respect to $\xi$. We noted at the end of Step 1 that the third and fourth terms of (3.12) are negligible. Besides (3.18) implies that we should take into account only the product of the first terms of the RHS’s of (3.12) and (3.14) for the asymptotic distribution of $\hat{\psi}(x) - \psi(x)$.

Noting that $\Omega_0^{-1}(0)$ is diagonal, we have only to consider

$$
\sum_{i=1}^{n} \int_{0}^{\tau} \left( \int_{0}^{x} \frac{1}{V(\xi)} \frac{X_i - \xi}{h} K_h(X_i - \xi) d\xi \right) M_i(dt).
$$

(3.20)

By putting $u = -(X_i - \xi)/h$ and using the Taylor expansion, the expression inside the parentheses is reduced to

$$
\int_{0}^{x} \frac{1}{V(\xi)} \frac{X_i - \xi}{h} K\left(\frac{X_i - \xi}{h}\right) d\xi
= - \int_{-X_i/h}^{(x-X_i)/h} V^{-1}(X_i) u K(u) du
- h \int_{-X_i/h}^{(x-X_i)/h} (V^{-1}(X_i + \theta_u u h))' u^2 K(u) du,
$$

(3.21)

where $0 < \theta_u < 1$.

We can easily show that the effect of the second term of the RHS of (3.21) on (3.20) is $O_p(n^{-1/2})$. The contribution of the first term to (3.20) is written as

$$
- \frac{1}{nh} \sum_{i=1}^{n} \int_{0}^{\tau} \left( \int_{-1}^{(x-X_i)/h} u K(u) du \right) V^{-1}(X_i) I(x - h < X_i < x + h) M_i(dt)
- \frac{1}{nh} \sum_{i=1}^{n} \int_{0}^{\tau} \left( \int_{-X_i/h}^{1} u K(u) du \right) V^{-1}(X_i) I(-h < X_i < h) M_i(dt).
$$

(3.22)

We used the symmetry of $K(\cdot)$ in (3.22).

Multiplying (3.22) by $(nh)^{1/2}$ and applying the CLT, we obtain the convergence in law in Theorem 2.

The bias of $\hat{\psi}(x) - \psi(x)$ comes from $U_2(\beta_0)$. The proof of the bias part of the theorem is easier than that of the convergence in law and is omitted. In fact the result on the bias is expected from Theorem 4 in Fan et al. (1997). Hence the proof of Theorem 2 is complete.

**Remark 1.** We give some comments on the cases of time-dependent covariates. Suppose that all the uniformity of convergence in the proof holds. Then we have Theorem 2 with $f(\xi)$ replaced with $f(\xi, t)$, where $f(\xi, t)$ is the density function of $X(t)$. Besides, we should define $g(\xi, t)$ by $E\{Y(t) \mid X(t) = \xi\}$. What we need for all the uniformity of convergence is essentially as follows:
(i) Some kind of the H"older continuity of $X(t)$. For example,
\[
\sup_{0 \leq s, t \leq \tau} \frac{|X(t) - X(s)|}{|t - s|^{\epsilon_1}} \leq W \quad \text{a.s. and } \quad \mathbb{E}\{W^{\epsilon_2}\} < \infty
\]
for some positive $\epsilon_1$ and $\epsilon_2$.

(ii) The existence of the bounded and continuous density of $X(T)$.

For example, we can treat Lemma 1 with the H"older continuity and the last term of the RHS of (3.16) is bounded from above by
\[
M \left( \frac{1}{n} \sum_{i=1}^{n} |R(\xi, T_i)| K_h(X_i(T_i) - \xi) + \int_{0}^{\tau} |R(\xi, t)| \frac{1}{n} \sum_{i=1}^{n} K_h(X_i(t) - \xi) \lambda_0(t) dt \right)
\]
for some positive constant $M$. We can handle $(1/n) \sum_{i=1}^{n} K_h(X_i(T_i) - \xi)$ and $(1/n) \sum_{i=1}^{n} K_h(X_i(t) - \xi)$ by the existence of the density of $X(T)$ and the H"older continuity, respectively.

Acknowledgements

The author thanks two anonymous reviewers for helpful comments.

REFERENCES


