CHAPTER 1

Empirical Plug-in Curve and Surface Estimates

Jiancheng Jiang  
Department of Probability & Statistics  
Peking University, Beijing 100871  
E-mail: Jiang@math.pku.edu.cn

Kjell Doksum  
Department of Statistics  
UC, Berkeley and UW, Madison  
E-mail: Doksum@stat.Berkeley.edu

We obtain estimates of a surface $\theta : \mathcal{R}^d \rightarrow \mathcal{R}$ by first finding the coefficient vector $\beta(x)$ that minimizes a distance between $\theta(x)$ and an approximating function $\theta(z-x; \beta)$ for $z$ in a neighborhood of a given point $x \in \mathcal{R}^d$, next expressing this $\beta(x)$ as a functional $\beta(x; G)$ for a surface $G : \mathcal{R}^q \rightarrow \mathcal{R}$ that admits an empirical estimate $\hat{G}(\cdot)$, and then using the empirical plug-in approach with $\beta(x) = \beta(x; G)$ and $\theta(x) = \theta(0, \beta(x))$. Examples of $G$'s are the multivariate distribution function and the integrated hazard function. With this approach the random part of the estimation error can be expressed in terms of empirical processes of the form $\int v(\cdot, x)d[\hat{G}(\cdot) - G(\cdot)]$, which facilitates asymptotic analysis. This approach gives simple derivations of old and new curve and surface estimates as well as their properties. It is particularly useful for providing simple and highly efficient estimates in boundary regions. We apply the approach to density estimation, univariate and multivariate; to hazard estimation; and to nonparametric regression.

KEY WORDS: Censoring; density; boundary adaptive; hazard rate; local polynomial; plug-in; regression.
1. Introduction

Let \( \theta(\cdot) : \mathcal{R}^d \to \mathcal{R} \) be a nonparametric surface and let \( \Theta = \{\theta(\cdot;\beta) : \mathcal{R}^d \to \mathcal{R}, \beta \in \mathcal{R}^q\} \) be a class of approximating functions. The class \( \Theta \) may or may not be a sieve. It is assumed that \( \Theta \) contains all constants.

Because \( \theta(\cdot) \) is unknown, estimators are often constructed by minimizing a distance between \( \theta(\cdot;\beta) \) and some data based function or value that is a proxy for \( \theta(\cdot) \). In the case of regression where \( \theta(x) = \mathbb{E}(Y|X = x) \), it is natural to use \( Y_i \) as a proxy of \( \theta(\cdot) \). However, in other cases, it is not clear what data based proxy for \( \theta(\cdot) \) one should use. For instance, for the cases where \( \theta(\cdot) \) are a density, a hazard function, and a hazard regression function, Jones (1993) (c.f. Fan and Gijbels (1996, Section 2.7.2)), Nielsen and Tanggaard (2001) and Jiang and Doksum (2002), respectively, used data based Dirac functions to construct proxies for \( \theta(\cdot) \).

However, Dirac functions do not exist in the ordinary sense, they only exist in the space of Schwartz distributions.

Here we consider a simple approach that only uses ordinary functions and gives the same solutions as the Dirac function approach when it applies. This approach also gives the same solutions as the one based on proxies that are ordinary functions. One version of the approach is simply to find the \( \beta = \beta(x) \) that locally, near \( x \), minimizes a distance based on \( \theta(z) - \theta(z - x; \beta) \). This \( \beta \) depends on the unknown \( \theta(\cdot) \), however, in many interesting cases it is possible to express this dependence through \( dG \), where \( G \) is a function that admits an empirical estimate \( \hat{G} \) such that \( \mathbb{E}G \) and integrals with respect to \( d\hat{G} \) exist in the ordinary sense. Examples of \( \hat{G} \) are the empirical distribution, the Kaplan Meier estimate and the Nelson-Aalen estimate. Let \( \hat{\beta}(x) \) be the estimate of \( \beta(x) \) obtained by plugging in the empirical estimate \( \hat{G} \) for \( G \), then \( \hat{\theta}(x) = \theta(0; \hat{\beta}(x)) \) is the empirical plug-in estimate of \( \theta(\cdot) \) evaluated at \( x \). Properties of \( \hat{\theta}(\cdot) \) can then be developed using well established properties of empirical estimates \( \hat{G} \). In the sections that follow we will apply this approach to old and new function estimation problems.

Here are some further details on how to carry out the approach: First note that we are faced with the usual overfitting problem because for each \( x \) there is a \( \beta \) such that \( \theta(x; \beta) = \theta(x) \). One way to avoid overfitting would be to introduce a roughness penalty. For instance, we could minimize the penalized distance

\[
\int \|\theta(x) - \theta(x;\beta)\|^2w(x)dx + \lambda \int \|\nabla^m\theta(x;\beta)\|^2dx,
\]

where \( \|\nabla^m\theta(x;\beta)\| \) is the Euclidean norm of the \( m \)th order partial deriv-
Plug-in Estimation

tives with respect to $x$ of $\theta(x; \beta)$, and $w(x)$ is a weight function. However, in this paper we will use kernel smoothing which insists on closeness of $\theta(z - x; \beta)$ to $\theta(z)$ for $z$ in a neighborhood of $x$. For a given $x$, we find the member $\theta(\cdot - x, \beta(x))$ of $\Theta$ that best approximates $\theta(\cdot)$ in a neighborhood of $x$ by setting

$$\beta(x) = \arg \min D_x(\theta(\cdot), \theta(\cdot - x; \beta)),$$

where

$$D_x(\theta(\cdot), \theta(\cdot - x; \beta)) = \int (\theta(z) - \theta(z - x; \beta))^2 K_H(z - x) w(z) dz.$$

Here $K_H(t) = K(H^{-1}t)/|H|$, $K(\cdot)$ is a $d$-variate probability density (the kernel) with unique mode 0 and $\int uK(u)du = 0$, $H$ is a nonsingular $d \times d$ matrix (the bandwidth matrix), and $w(\cdot)$ is a nonnegative weight function which is continuous and nonzero at $x$. The best local approximation to $\theta(x)$ in $\Theta$ is $\theta(0, \beta(x))$.

A minimizer $\beta(x)$ of $D_x$ exists if

$$\lim_{||\beta|| \to \infty} \theta(x; \beta) = \infty,$$

where $|| \cdot ||$ is the Euclidean norm. If further $\theta(x; \beta)$ is differentiable in $\beta$, then $\beta = \beta(x)$ must satisfy the equations (for $j = 1, \ldots, q$)

$$\int \theta(z) K_H(z - x) w(z) \nabla_j \theta(z - x; \beta) dz$$

$$= \int \theta(z - x; \beta) \nabla_j \theta(z - x; \beta) K_H(z - x) w(z) dz, \quad (1)$$

where $\nabla_j$ denotes the partial derivative with respect to $\beta_j$. It turns out that in many interesting cases we can choose $w(\cdot)$ so that the solution $\beta(x)$ to these equations can be expressed explicitly in terms of integrals of the form

$$\int v(z) K_H(z - x) dG(z),$$

where $v(z)$ is a weight function and $G$ is a function that admits a nonparametric empirical estimate $\hat{G}$. Thus by plugging in $d\hat{G}$ for $dG$, we obtain an empirical plug-in estimate $\hat{\beta}(x)$ of $\beta(x)$ and $\hat{\theta}(x) = \theta(0; \hat{\beta}(x))$ is the empirical plug-in estimate of $\theta(x)$. In particular, if $\theta(x; \beta)$ is a polynomial of order $p$, then $\partial^k \theta(z; \beta) / \partial z^k |_{z=0, \beta=\beta(x)}$ is the empirical plug-in estimate of $\partial^k \theta(x) / \partial x^k$. 
Whenever there is a natural data based proxy $\hat{\theta}(x)$ for $\theta(\cdot)$, e.g. $\theta(\cdot)$ admits a nonparametric empirical estimate $\hat{\theta}(x)$, then we can choose $\tilde{\beta} = \tilde{\beta}(x)$ to minimize

$$\rho_x(\hat{\theta}(\cdot), \theta(\cdot - x, \beta)) = \int [\hat{\theta}(z) - \theta(z - x; \beta)]^2 K_H(z - x) w(z) dz. \quad (2)$$

This $\tilde{\beta}(x)$ will typically coincide with $\hat{\beta}(x)$. Thus the empirical plug-in approach is general in the sense that it is applicable when a data based proxy for $\theta(\cdot)$ is available as well as when no such proxy exists in the ordinary sense.

With the empirical plug-in approach, it is convenient to analyze the asymptotic properties of the resulting estimates $\hat{\beta}(x)$, because the random part of the estimation error can be expressed in terms of empirical processes of the form $\int v(z) K_H(z - x) [\hat{G}(\cdot) - G(\cdot)]$, and the bias of the estimates can be easily obtained from the difference between $\int v(z) K_H(z - x) dG(\cdot)$ and the estimand $\theta(\cdot)$ and its derivatives. We will show that the estimates are boundary adaptive and highly efficient.

2. Plug-in Density Estimation

2.1. The one dimension case

We assume that $X_1, \ldots, X_n$ are i.i.d. with distribution function $F$ and density $f = F'$, $X_i \in \mathbb{R}$. Let

$$f(z - x; \beta) = \beta_0 + \sum_{j=1}^{p} \beta_j (z - x)^j$$

be a local polynomial approximation to $f(x)$ for $z$ near $x$. Now $q = p + 1$, $K_b(\cdot) = b^{-1} K(./b)$, where $K$ is a unimodal density with $\int u K(u) du = 0$, and

$$D_x(f(\cdot), f(\cdot - x; \beta)) = \int [f(z) - f(z - x; \beta)]^2 K_b(z - x) w(z) dz.$$

The minimizer $\beta(x)$ of $D_x$ satisfies the equations (for $\ell = 0, \ldots, p$)

$$\int (z - x)^\ell K_b(z - x) w(z) dF(z) = \sum_{j=0}^{p} \beta_j \int (z - x)^{j+\ell} K_b(z - x) w(z) dz.$$

That is, $\beta(x)$ can be expressed as a functional of $F(x)$

$$\mathbf{B}\beta(x) = \mathbf{C}^{-1} \mathbf{a}(F), \quad (3)$$
where \( B = \text{diag}(1, b, \cdots, b^p) \), \( C = (c_{jk}) \), \( a(F) = (a_0(F), \cdots, a_p(F))^T \),
\[
c_{jk} = \int \left( \frac{z - x}{b} \right)^{j+\ell} K_b(z - x)w(z)dz,
\]
\[
a_\ell(F) = \int \left( \frac{z - x}{b} \right)^\ell K_b(z - x)w(z)dF(z), \quad j, \ell = 0, \cdots, p.
\]
Replacing \( dF \) with \( d\hat{F} \) where \( \hat{F} \) is the empirical distribution gives us
\[
\hat{\beta}(x) = B^{-1}C^{-1}a(\hat{F}),
\]
which estimates the density function \( f(x) \) and its derivatives \( f(x) = (f(x), \cdots, f^{(p)}(x)/p)^T \). The empirical plug-in estimate of \( f(k)(x) \) (for \( k = 0, \cdots, p \)) is \( \hat{f}(k)(x) = k! \partial^k f(u; \beta)/\partial u^k \big|_{u=\hat{\beta}(x)} = k!\hat{\beta}_k \).

Assume that the supports of \( f(x) \) and \( K(u) \) are \([0, 1]\) and \([-1, 1]\), respectively. Let \( c = \min(1, \max(0, x/b)) \) and \( d = \max(0, \min(1, (1 - x)/b)) \) for \( x \in [0, 1] \). Let \( s_\ell = \int_{-c}^d K(u)u^\ell du, \quad \nu_\ell = \int_{-c}^d u^K(u)du, \quad c_p = (s_{p+1}, \cdots, s_{2p+1})^T, \quad c_{p+1} = (s_{p+2}, \cdots, s_{2p+2})^T, \quad S = (s_{i+j}) \) and \( V^* = (v_{i+j}) \) \((0 \leq i \leq p; 0 \leq j \leq p)\). Then we have the following theorems which describe the asymptotic properties of the empirical plug-in estimates for the density and its derivatives.

**Theorem 2.1:** Under conditions (A1)-(A4) in Appendix I, for \( x \in [0, 1] \)
\[
\sqrt{nB} \{ B[\hat{\beta}(x) - f(x)] - \frac{f^{(p+1)}(x)b^{p+1}}{(p+1)!} S^{-1}c_p
- \frac{f^{(p+2)}(x)b^{p+2}}{(p+2)!} S^{-1}c_{p+1} + o(b^{p+2}) \} \xrightarrow{\mathcal{L}} \mathcal{N}(0, S^{-1}V^*S^{-1}f(x)).
\]

**Remark 2.1:** From the proof of Theorem 2.1, one sees that, when estimating a density function which is a polynomial of order \( p \) on an interval, the finite sample bias of the plug-in estimators on the interval is zero. This contrasts with the kernel density estimates based on higher order kernels (Müller, 1984), for which the respective zero bias only holds true asymptotically.

Note that when \( x \) is an interior point (i.e. \( x \in (b, 1 - b) \)) and all odd-order moments of \( K(\cdot) \) vanish, that is, \( \int_1^b u^K(u)du = 0 \) for \( \ell \) odd, some of the entries in \( S^{-1}c_p \) are zero and Theorem 2.1 is not very informative. We now take a closer look at this issue. Let \( e_k = (0, \cdots, 1, \cdots, 0)^T \), where
\( e_k \) has one in the \((k+1)\)-th component and zeros in the others. Then from Theorem 2.1

**Theorem 2.2:** Suppose \( K(\cdot) \) is symmetric. Under conditions (A1)-(A4) in Appendix I, for an interior point \( x \in (b, 1-b) \)

(i) For \( p-k \) odd

\[
\sqrt{n}b^{2k+1} \left\{ \frac{1}{k!} \left[ f^{(k)}(x) - f^{(k)}(x) \right] - \frac{f^{(p+1)}(x)b^{p+1-k}}{(p+1)!} e_k^T S^{-1} c_p(1 + o(1)) \right\} \\
\xrightarrow{\mathcal{L}} N \left( 0, e_k^T S^{-1} V S^{-1} e_k f(x) \right).
\]

(ii) For \( p-k \) even

\[
\sqrt{n}b^{2k+1} \left\{ \frac{1}{k!} \left[ f^{(k)}(x) - f^{(k)}(x) \right] - \frac{f^{(p+2)}(x)b^{p+2-k}}{(p+2)!} e_k^T S^{-1} c_{p+1}(1 + o(1)) \right\} \\
\xrightarrow{\mathcal{L}} N \left( 0, e_k^T S^{-1} V S^{-1} e_k f(x) \right).
\]

**Remark 2.2:** From Theorem 2.2, the asymptotic mean squared error \((AMSE_k)\) for estimating \( f^{(k)}(x)/k! \) \((k = 0, \ldots, p)\) is

\[
b^{2(p+1-k)} [e_k^T S^{-1} c_p f^{(p+1)}(x)/(p+1)!]^2 + \frac{1}{nb^{2k+1}} e_k^T S^{-1} V S^{-1} e_k f(x),
\]

for \( p-k \) odd; and

\[
b^{2(p+2-k)} [e_k^T S^{-1} c_{p+1} f^{(p+2)}(x)/(p+2)!]^2 + \frac{1}{nb^{2k+1}} e_k^T S^{-1} V S^{-1} e_k f(x),
\]

for \( p-k \) even. Therefore, the optimal local bandwidth \( b_{k,\text{opt}}(x) \) for estimating the \( k \)-th derivative of \( f(x) \) at \( x \), in the sense of minimizing \( AMSE_k(b, x) \), is

\[
n^{-\frac{1}{2p+3}} \left( \frac{[(p+1)!]^2 e_k^T S^{-1} V S^{-1} e_k f(x)}{2(p+1-k)[f^{(p+1)}(x)]^2(e_k^T S^{-1} c_p)^2} \right)^{\frac{1}{2p+3}}, \tag{5}
\]

if \( p-k \) is odd; and

\[
n^{-\frac{1}{2p+3}} \left( \frac{[(p+2)!]^2 e_k^T S^{-1} V S^{-1} e_k f(x)}{2(p+2-k)[f^{(p+2)}(x)]^2(e_k^T S^{-1} c_{p+1})^2} \right)^{\frac{1}{2p+3}}, \tag{6}
\]

if \( p-k \) is even.
In parallel to Theorem 2.2, consider the estimation of \( f^{(k)}(x) \) at the left boundary point \( x \in [0, b) \). Since \( \mathbf{e}_k^T \mathbf{S}_d^{-1} \mathbf{e}_p \) does not vanish, we have the following result from Theorem 2.1:

**Theorem 2.3:** Suppose \( K(\cdot) \) is symmetric. Under conditions (A1)-(A4) in Appendix I, for \( x \in [0, b) \)

\[
\sqrt{n}b^{2k+1} \left\{ \frac{1}{k!} \int f^{(k)}(x) - f^{(k)}(x) dx - \frac{1}{(p+1)!} f^{(p+1)}(0)b^{p+1-k} \mathbf{e}_k^T \mathbf{S}_d^{-1} \mathbf{e}_p + o(b^{p+1-k}) \right\} \overset{L}{\rightarrow} \mathcal{N} \left( 0, \mathbf{e}_k^T \mathbf{S}_d^{-1} \mathbf{V}_d^{-1} \mathbf{e}_k f(0) \right).
\]

**Remark 2.3:** From Theorems 2.2 and 2.3, the plug-in estimate of \( f^{(k)}(\cdot) \) (for \( p - k \) odd) is boundary adaptive in the sense that it automatically achieves the same convergence rate \( O(n^{-\frac{2(k+1)}{2p+2}}) \) on boundary points as in interior regions if a symmetric kernel and the optimal bandwidths \( b_{k,opt} = O(n^{-\frac{1}{2p+2}}} \) in (5) are employed. For \( p - k \) even, the estimator is also consistent (even if \( p = 0 \) is employed), but the bias at the boundary is of order lower than at interior points if the optimal bandwidths \( b_{k,opt} \) in (6) is used. This is similar to the case of the well-known local polynomial regression.

### 2.2. The multivariate case

Let \( \mathbf{X}_1, \cdots, \mathbf{X}_n \) be i.i.d. as \( \mathbf{X} \sim F \), where \( \mathbf{X} \in \mathbb{R}^d \). The empirical distribution function is \( \hat{F}(x) = n^{-1} \sum_{i=1}^n 1[\mathbf{X}_i \leq x] \), where \( \mathbf{X} \leq \mathbf{x} \) is componentwise inequality, \( \mathbf{x} \in \mathbb{R}^d \). We let \( K(\mathbf{u}) \) denote a \( d \)-variate kernel which is a \( d \)-variate density function with the marginals of \( K \) are all equal, say to \( K_1(u) \), and \( \int u_i u_j K(\mathbf{u}) d\mathbf{u} = 1[i = j] \mu_2(K) \) where \( \mu_2(K) = \int u_i^2 K_1(u) du \) and \( u_i \) is the \( i \)th component of \( \mathbf{u} \). Let \( \mathbf{H} = \text{diag}(h_1, \cdots, h_d) \) with \( h_j > 0 \) and set \( K_{\mathbf{H}}(\mathbf{u}) = |\mathbf{H}|^{-\frac{1}{2}} K(\mathbf{H}^{-1} \mathbf{u}) \). Assume that the support \( \Omega = \{ \mathbf{x} : f(\mathbf{x}) > 0 \} \) of \( f \) is compact. Let \( \alpha(\mathbf{x}) \) and \( \beta(\mathbf{x}) = (\beta_1(\mathbf{x}), \cdots, \beta_d(\mathbf{x}))^T \) be the minimizers of

\[
D_x(f(\cdot), f(\cdot - \mathbf{x}; \alpha, \beta)) \equiv \int_\Omega \{ f(t) - [\alpha + \beta^T (t - \mathbf{x})] \}^2 K_{\mathbf{H}}(t - \mathbf{x}) dt \quad (7)
\]

and define the boundary correcting kernel estimate as \( \hat{f}(\mathbf{x}) = \hat{\alpha}(\mathbf{x}) \) and \( \nabla \hat{f}(\mathbf{x}) = \hat{\beta}(\mathbf{x}) \), where \( (\hat{\alpha}(\mathbf{x}), \hat{\beta}^T(\mathbf{x})) \) are the empirical plug-in estimates and \( \nabla f(\mathbf{x}) \) denotes the column vector of first-order partial derivatives of
f(x). That is, \((\hat{\alpha}(x), \hat{\beta}^T(x))\) is the solution to the following linear equation

\[ S_{x,H} \cdot \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = n^{-1} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ X_i - x \end{bmatrix} K_H(X_i - x), \tag{8} \]

where

\[ S_{x,H} = \begin{bmatrix} \int_{\Omega} K_H(t - x)dt & \int_{\Omega} (t - x)^T K_H(t - x)dt \\ \int_{\Omega} (t - x)K_H(t - x)dt & \int_{\Omega} (t - x)^2 K_H(t - x)dt \end{bmatrix}. \]

It is worthwhile to investigate the boundary behavior of the empirical plug-in estimates particular at the multivariate case. We will use the convenient mathematical formulation of the edge effect problem given by Ruppert and Wand (1994). Let \(\varepsilon_{x,H} = \{z: H^{-1}(x - z) \in \text{supp}(f)\}\) be the support of \(K_H(x - \cdot)\). We call \(x\) interior if \(\varepsilon_{x,H} \subset \text{supp}(f)\). Otherwise, \(x\) is called a boundary point. Let \(D_{x,H} = \{z: (x + Hz) \in \text{supp}(f) \cap \text{supp}(K)\}\). Then \(D_{x,H} = \text{supp}(K)\) if and only if \(x\) is an interior point. When \(\max |h_j| \to 0\) as \(n \to \infty\), a fixed point \(x\) in the interior of \(\text{supp}(f)\) is an interior point for large \(n\). Denote by \(x_\partial\) a point on the boundary of \(\text{supp}(f)\). Assume that there is a convex set \(C\) with nonnull interior and containing \(x_\partial\) such that

\[ \inf_{x \in C} f(x) > 0. \tag{9} \]

Simple algebra gives that

\[ S_{x,H} = ANA, \]

where

\[ A = \begin{bmatrix} 1 & 0 \\ 0 & H \end{bmatrix} \quad \text{and} \quad N = \int_{D_{x,H}} \begin{bmatrix} 1 \\ u \end{bmatrix} [1 \ u^T] K(u)du. \]

The solution to the linear equation (8) is then

\[ (\hat{\alpha}(x), \hat{\beta}^T(x))^T = A^{-1}N_x^{-1}A^{-1}S_n(x), \tag{10} \]

where

\[ A = \begin{bmatrix} 1 & 0 \\ 0 & H \end{bmatrix}, \]

\[ N_x = \int_{D_{x,H}} \begin{bmatrix} 1 \\ u \end{bmatrix} [1 \ u^T] K(u)du, \]

and

\[ S_n(x) = n^{-1} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ X_i - x \end{bmatrix} K_H(X_i - x). \]
Plug-in Estimation

Note that for an interior point \( x \), \( D_{x,H} = \text{supp}(K) \) and if \( K(\cdot) \) is a \( d \)-variate density with \( \int K(u)du = 1, \int uK(u)du = 0 \), then

\[
N_x = \begin{bmatrix} 1 & 0 \\ 0 & \int u^T K(u)du \end{bmatrix}
\]

and the solution to (8) is

\[
\hat{f}(x) = n^{-1} \sum_{i=1}^{n} K(H(X_i - x)),
\]

which is the usual multivariate kernel density estimate. However, for a boundary point, \( x = x_\partial + Hc \) say (for a fixed \( c \) in \( \text{supp}(K) \)), the plug-in estimate is different but automatically boundary-corrected, as shown in the following theorem.

Let \( \mathcal{H}_f(x) \) be the \( d \times d \) Hessian matrix of \( f(x) \).

**Theorem 2.4:** Suppose that (9) holds. Then under conditions (B1)-(B3) in Appendix I, for any \( x \in \Omega \)

\[
\sqrt{n|H|} \left[ \begin{bmatrix} \hat{a}(x) - f(x) \\ H[\hat{a}(x) - \nabla f(x)] \end{bmatrix} - \frac{1}{2} N_x^{-1} C_x (1 + o(1)) \right] \xrightarrow{L} \mathcal{N}(0, N_x^{-1} T_x N_x^{-1} f(x)),
\]

where

\[
C_x = \int_{D_{x,H}} \begin{bmatrix} 1 \\ u \end{bmatrix} K(u)u^T H \mathcal{H}_f(x) Hu du,
\]

\[
T_x = \int_{D_{x,H}} \begin{bmatrix} 1 \\ u \end{bmatrix} [1, u^T] K^2(u) du.
\]

**Corollary 2.1:** For a symmetric density kernel \( K(\cdot) \) and an interior point \( x \),

\[
\sqrt{n|H|} \left[ \hat{f}(x) - f(x) - \frac{1}{2} \mu_2(K) tr \{ H \mathcal{H}_f(x) H \} (1 + o(1)) \right] \xrightarrow{L} \mathcal{N}(0, f(x) \int K^2(u)du).
\]

From the theorem, we can see that the empirical plug-in estimate of the multivariate density is automatically boundary corrected. The optimal bandwidth can be developed in the same way as in the one dimensional case.
3. Hazard Estimation with Censored Data

Let $T_1, \cdots, T_n$ be i.i.d. lifetimes distributed as $T$ where $T \geq 0$, $T$ has density $f$, distribution function $F$, hazard function $\lambda(x) = \Lambda'(x) = f(x)/(1 - F(x))$ and integrated hazard function $\Lambda(x) = -\log(1 - F(x))$. We are interested in estimating $\lambda(\cdot)$ based on $X_i = \min(T_i, C_i)$ and $\delta_i$, where $C_1, \cdots, C_n$ are i.i.d. censoring times with distribution function $G$, $\delta_i = 1[T_i \geq C_i] = 1[T_i = C_i]$, $i = 1, \cdots, n$, and the $C$’s are independent of the $X$’s.

Let $L$ denote the distribution function of $X_i$, then $L = \bar{F} \bar{G}$, where for any distribution function $H$, $\bar{H} = 1 - H$ is the corresponding survival function. We consider the problem of estimating $\lambda^{(k)}(x)$ on the interval $[0, T]$ (for $k = 0, 1, \cdots$), where $T^* < \inf \{x : L(x) = 1\}$.

For a given kernel $K(\cdot)$, let $c = \min(1, \max(0, x/b))$, $s_{t,c} = \int_{-c}^{t} K(u)u' \, du$, $v_{t,c} = \int_{-c}^{t} u' K^2(u) \, du$, $S_c = (s_{i+j,c})$ and $V_c^* = (v_{i+j,c})$, $c_{p,c} = (s_{p+1,c}, \cdots, s_{2p+1,c})^T$, $c_{p+1,c} = (s_{p+2,c}, \cdots, s_{2p+2,c})^T$, $(0 \leq i \leq p; 0 \leq j \leq p; 0 \leq \ell \leq 2p + 2)$.

3.1. The “classic” approach

Consider finding $\beta(x) = \beta(x, \Lambda)$ that minimizes

$$\min_{\beta} \int_{u \geq 0} K_b(u-x)|\lambda(u) - \sum_{j=0}^{p} \beta_j (u-x)^j|^2 \, du. \quad \text{(11)}$$

By Taylor expansion, the solution of the optimization problem, denoted by $\beta(x) = (\beta_0, \cdots, \beta_p)^T$, approximates $a(x) = (\lambda(x), \cdots, \lambda^{(p)}(x)/p!)^T$. Our basic empirical estimating equations can be written as

$$\int_{u \geq 0} K_b(u-x)(u-x)^\ell \, d\Lambda(u) = \sum_{j=0}^{p} \beta_j \int_{u \geq 0} (u-x)^{j+\ell} \, K_b(u-x) \, du.$$ 

We will use the Nelson-Aalen’s estimator of $\Lambda(x)$:

$$\hat{\Lambda}(x) = \int_{0}^{x} (1 - L_n(u))^{-1} dL_{1n}(u) = \sum_{i : X_{(i)} \leq x} \delta_{(i)}/(n - i + 1),$$

where $X_{(i)}$ are the order statistics of $X_i$, and $\delta_{(i)}$ is the concomitant of $X_{(i)}$, i.e. $\delta_{(i)}$ and $X_{(i)}$ are from the same observation. It follows that the empirical plug-in estimate can be written as

$$\hat{\beta}(x) = \beta(x, \hat{\Lambda}) = H^{-1} S_e^{-1} \cdot T(x, \hat{\Lambda}), \quad \text{(12)}$$
Plug-in Estimation

where \( T(x, \hat{\Lambda}) = (T_0(x, \hat{\Lambda}), \ldots, T_p(x, \hat{\Lambda}))^T \), and for \( \ell = 0, \ldots, p \)

\[
T_\ell(x, \hat{\Lambda}) = \int_{u \geq 0} K_b(u - x)(\frac{u - x}{b})^\ell d\hat{\Lambda}(u)
\]

\[
= \sum_{i=1}^{n} K_b(X(i) - x) \left( \frac{X(i) - x}{b} \right)^\ell \frac{\delta(i)}{n - i + 1}.
\]

The plug-in estimator \( \hat{\beta}(x) \) is exactly the same as the locally polynomial estimator of Jiang and Doksum (2002). Its asymptotic distribution can now be derived (see Jiang and Doksum 2002) by using well known properties of the process \( \sqrt{n}[\hat{\Lambda}(x) - \Lambda(x)] \). In particular,

**Theorem 3.1:** If \( K(\cdot) \) is symmetric, and if \( p - k \) is odd, then under conditions (C1) – (C4) in Appendix I,

(i) If \( x \) is an interior point, i.e. \( x \in (b, T^*) \),

\[
\sqrt{n} b^{2k+1} \left\{ \frac{1}{k!} (\lambda^{(k)}(x) - \lambda^{(k)}(\hat{x})) - \frac{\lambda^{(p+1)}(0)b^{p+1-k}}{(p+1)!} e_k^T S^{-1} c_p(1 + o(1)) \right\}
\]

\[
\xrightarrow{L} \mathcal{N}(0, e_k^T S^{-1} V^* S^{-1} e_k \lambda(0)/\bar{L}(x)),
\]

where \( S, V^* \) and \( c_p \) are the same as \( S_c, V_c^* \) and \( c_{p,c} \) but with \( c = 1 \)

(ii) If \( x \) is a boundary point, \( x = cb \) \((0 \leq c < 1)\), then

\[
\sqrt{n} b^{2k+1} \left\{ \frac{1}{k!} (\lambda^{(k)}(x) - \lambda^{(k)}(\hat{x})) - \frac{\lambda^{(p+1)}(0)b^{p+1-k}}{(p+1)!} e_k^T S_c^{-1} c_{p,c}(1 + o(1)) \right\}
\]

\[
\xrightarrow{L} \mathcal{N}(0, e_k^T S_c^{-1} V_c^* S_c^{-1} e_k \lambda(0)/\bar{L}(0)).
\]

**3.2. The counting process approach**

The counting process approach, due to Aalen (1975, 1978), allows for more general frameworks in hazard estimation. For \( n \) independently observed cases, let \( N_i(t) \) be 1 if the \( i \)th case failed in the interval \((0, t] \) and 0 otherwise; and let \( Y_i(t) \) be 1 if the \( i \)th case is still at risk at time \( t \) and 0 otherwise. In our previous notation,

\[
N_i(t) = 1(X_i \leq t, \delta_i = 1), \quad Y_i(t) = 1(X_i > t),
\]

where \((N_i(t), Y_i(t)): i = 1, \ldots, n\) are i.i.d.. See page 137 of Andersen, Borgan, Gill and Keiding (1993). In this section we compare our approach
to that of Nielsen (1998) and Nielsen and Tanggaard (2001), who proposed the estimate

$$\hat{\lambda}(t) = \arg \min_{\Theta} \sum_{i=1}^{n} \int_{0}^{T} |\Delta N_i(s) - \beta_0 - \beta_1(t-s)|^2 K_b(t-s) W(s) Y_i(s) ds,$$

(13)

where $\Delta N_i(s)$ is the Dirac “derivative” of $N_i(s)$ defined by the property

$$\int v(s) \Delta N_i(s) ds = \int v(s) dN_i(s)$$

for every integrable function $v(\cdot)$. Our empirical plug-in approach would be to replace $N_i(s)$ by its expected value $L_1(s) = E[N_i(s)]$, replace $\Delta N_i(s)$ with the subdistribution density $\ell_1(s) = L'_1(s)$ for the uncensored observations, replace $\beta_0 + \beta_1(t-s)$ by $\sum_{j=0}^{p} \beta_j(s-t)^j$, solve the new (13) for $\beta$ in terms of $dL_1(\cdot)$, and to replace $dL_1(t)$ with $dN^{(n)}(t) = n^{-1} \sum_{i=1}^{n} dN_i(t)$ to get $\ell_1(t) = \hat{\beta}_0(t)$. However,

$$L_1(t) = E[N_i(t)] = \int_{0}^{t} G(v)f(v) dv,$$

$$\ell_1(t) = \hat{G}(t)f(t) \neq \lambda(t) = f(t)/\bar{F}(t).$$

Thus our approach would give an estimate of $\ell_1(t) = \hat{G}(t)f(t) \neq \lambda(t)$. To remedy this, note that $\hat{L}(t) = P(X_i \geq t) = \bar{F}(t)/\bar{G}(t)$ and $\lambda(t) = \ell(t)/\hat{L}(t)$, if we set $\hat{\lambda}_1(t) = \hat{\ell}_1(t)/\hat{L}(t)$, where $\hat{L}(t) = n^{-1} \sum_{i=1}^{n} 1(X_i \geq t)$, then $\hat{\lambda}_1(t)$ provides an estimate of $\lambda(t)$.

Another empirical plug-in approach would be to replace (13) with

$$\hat{\ell}_1(t) = \arg \min \int [\ell_1(s) - \sum_{j=0}^{p} \beta_j(s-t)^j]^2 K_b(s-t)v(s) ds,$$

(14)

where $v(s)$ is a weight function which can be a random process, e.g. $v(s) = W(s)Y^{(n)}(s)$ where $Y^{(n)}(s) = \sum_{i=1}^{n} Y_i(s)$ and $W(s)$ is as in (13). Note that (14) can be solved explicitly for $\beta = \beta(t,dL_1)$ and the empirical plug-in estimate of $\lambda(t)$ is

$$\hat{\lambda}_2(t) = \hat{\beta}_0(t,dN^{(n)}(t))/\hat{L}(t).$$

If $v(s) = W(s)Y^{(n)}(s)$, then some algebra gives $\hat{\lambda}_1(t) = \hat{\lambda}_2(t)$.

As in Nielsen and Tanggaard (2001), we assume that the weight function $W(s)$ and the exposure have local limits $\gamma(s)$ and $w(s)$ which are continuous
at time $t$, i.e.
\[
\sup_{s \in \Delta_{t,b}} |Y^{(n)}(s)/n - \gamma(s)| \to 0, \tag{15}
\]
\[
\sup_{s \in \Delta_{t,b}} |W(s) - w(s)| \to 0, \tag{16}
\]
where $\Delta_{t,b} = [t-10b, t + 10b]$, and $\gamma(s)$ and $w(s)$ are continuous at time $t$.

**Theorem 3.2:** Under conditions (C1)-(C4) in Appendix I and (15)-(16), for $p$ odd, $i = 1, 2$ and $t \in [0, T^*]$.
\[
\sqrt{n}b[\hat{\lambda}(t) - \lambda(t) - \frac{\ell^{(p+1)}(t) (p+1)! L(t)}{(p+1)! L(t)} e^T S^{-1} c_{p,c} (1 + o(1))] \to N(0, \sigma^2(t)),
\]
where $\sigma^2(t) = e^T S^{-1} V S^{-1} e \lambda(t)/L(t)$.

The plug-in estimate has the same asymptotic variance as that in Theorem 3.1, but their asymptotic bias are different although they are of the same order. Note that the above plug-in estimates are not smooth since $\hat{L}(t)$ is empirical estimate which is not continuous. To get a smooth $\hat{\lambda}(t)$, one can employ a smooth technique for estimating $\hat{L}(t)$ such as the one in Jiang, et. al (2002). Since the resulted $\hat{L}(t)$ is $\sqrt{n}$-consistent, the asymptotic normality remains the same as before.

**4. Plug-in Estimation of the Regression Function**

Given i.i.d observations $\{X_i, Y_i\}_{i=1}^n$ of $(X,Y)$, where $(X,Y)$ have density $f(\cdot, \cdot)$ and distribution function $F(\cdot, \cdot)$, consider estimation of the regression function $m(x) = E[Y|X=x]$. In the following, we will illustrate that our empirical plug-in approach reduces to the kernel smoother, the local linear smoother, and others.

**4.1. Local polynomial regression as a special case of empirical plug-in estimation**

Let $m(z - x; \beta) = \beta_0 + \sum_{j=1}^p \beta_j (z - x)^j$. Then a familiar approach is to minimize
\[
\sum_{i=1}^n \left[ \hat{m}(X_i) - m(X_i - x; \beta) \right]^2 K_h(X_i - x). \tag{17}
\]
We can choose $\hat{m}(X_i) = Y_i$ (i.e. Fan and Gijbels 1996, page 58), which is the same as the empirical plug-in estimate obtained by computing $E_{P}(Y|X_i) =$
\( Y_i \), where \( \hat{P} \) is the empirical probability which assigns probability \( 1/n \) to each \( (X_i, Y_i) \), \( i = 1 \cdots n \). When \( \hat{m}(X_i) = Y_i \), the solution to (17) is (e.g. Fan and Gijbels 1996, page 59),

\[
\hat{\beta}(x) = (X^T W K X)^{-1} X^T W Y,
\]

where \( X \) is a matrix with \((i, j)\)th element \((X_i - x)^j \) for \( 1 \leq i \leq n \) and \( 0 \leq j \leq p \), \( Y = (Y_1, \cdots, Y_n)^T \), and \( W = \text{diag}(K_h(X_1 - x), \cdots, K_h(X_n - x)) \).

The empirical plug-in approach yields the same \( \hat{\beta}(x) \) as a special case. To see this set \( w(z) = f_1(z)v(z) \), where \( f_1(z) \) is the marginal density of \( X \), and \( v(z) \geq 0 \) is a weight function which is continuous and positive at \( x \). Then the left hand side of (1) becomes

\[
\int \left[ \int \frac{yf(z, y)}{f_1(z)} \, dy \right] K_h(z - x)f_1(z)v(z)\nabla_j m(z - x; \beta) \, dz = \int yK_h(z - x)v(z)\nabla_j m(z - x; \beta) \, dF(z, y),
\]

where \( \nabla_j m(z - x; \beta) \) denotes the partial derivative of \( m(z - x; \beta) \) with respect to \( \beta_j \). Let \( F_1 \) denote the marginal distribution function of \( X \), then the right hand side of (1) becomes

\[
\int m(z - x; \beta)\nabla_j m(z - x; \beta) K_h(z - x)v(z) \, dF_1(z).
\]

If we choose \( v(z) = \text{constant} > 0 \), replace \( F \) and \( F_1 \) with their empirical estimates, we find that (18) solves (1).

### 4.2. A different plug-in estimate of regression

We define \( G(x) = \int yf(x, y) \, dy \); then \( m(x) = G(x)/f_1(x) \). In Section 2, we defined the plug-in estimate \( f_1(\cdot) \). Similarly, we can define a plug-in estimate of \( G(\cdot) \). Assume that the support of \( f_1(x) \) is compact, \([0, 1]\) say. Let us consider the following optimization problem: for \( p = 0, 1, 2, \cdots \),

\[
\min_{\beta} \int_0^1 [G(u) - \sum_{j=0}^p \beta_j (u - x)^j] K_h(u - x) \, du.
\]

By Taylor expansion, the solution of the optimization problem, denoted by \( \beta(x) = (\beta_0, \cdots, \beta_p)^T \), approximates \( (G(x), \cdots, G^{(p)}(x)/p!)^T \). The left hand side of (1) becomes

\[
\int K_h(u - x) (u - x)^T y \, dF(x, y)
\]
and the plug-in estimate $\hat{\beta}(x)$ is the solution to the linear equations: for $\ell = 0, \cdots, p$,
\[
n^{-1} \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^\ell Y_i = \sum_{j=0}^{p} \beta_j \int_{0}^{1} (u - x)^{j+\ell} K_h(u - x) du.
\] (20)

Let $c = \min(1, \max(0, x/h))$ and $d = \max(0, \min(1, (1-x)/h))$ for $x \in [0, 1]$. Then by a change of variable (20) can be rewritten as
\[
n^{-1} \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^\ell Y_i = \sum_{j=0}^{p} \beta_j \int_{-c}^{d} (th)^{j+\ell} K(t) dt.
\] (21)

It follows that the plug-in estimator $\hat{\beta}(x)$ admits the following form
\[
H \hat{\beta}(x) = S(x)^{-1} \cdot S_n(x),
\] (22)
where $H = \text{diag}(1, h, \cdots, h^p)$, $S(x) = (s_{i+j})$ is a $(p+1) \times (p+1)$ matrix with $s_{i+j} = \int_{-c}^{d} K(u) u^{i+j} du$ for $i, j = 0, \cdots, \ell$, $S_n(x) = (S_{n0}(x), \cdots, S_{np}(x))^T$, and
\[
S_{n\ell}(x) = \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^\ell Y_i, \text{ for } \ell = 0, \cdots, p.
\] (23)

Similarly, we have the following expression for $\hat{f}_1(x) \equiv (\hat{f}_1, \cdots, \hat{f}_1^{(p)}/p!)^T$ (see Section 2.1):
\[
H \hat{f}_1(x) = S(x)^{-1} \cdot T_n(x),
\] (24)
where
\[
T_{n\ell}(x) = \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^\ell, \text{ for } \ell = 0, \cdots, p.
\] (25)

Then
\[
\hat{m}(x) = \hat{G}(x)/\hat{f}_1(x) = \frac{e_1^T S(x)^{-1} \cdot S_n(x)}{e_1^T S(x)^{-1} \cdot T_n(x)},
\]
where $e_1^T = (1, 0, \cdots, 0)$. In particular, when $p = 0$ is employed, the estimator is
\[
\hat{m}(x) = S_{n0}(x)/T_{n0}(x) = \frac{\sum_{i=1}^{n} K_h(X_i - x) Y_i}{\sum_{i=1}^{n} K_h(X_i - x)},
\]
which is just the Nadaraya-Watson ($N$-$W$) estimate (Nadaraya 1964; Watson 1964). For $p = 1$, we have
\[
\hat{m}(x) = \frac{s_2 S_{n0}(x) - s_1 S_{n1}(x)}{s_2 T_{n0}(x) - s_1 T_{n1}(x)}.
\] (26)
If a symmetric kernel \( K(\cdot) \) is employed, then for an interior point \( x = 0 \) (i.e. \( x \in [h, 1-h] \)), \( s_1 = 0 \) and \( \hat{m}(x) = S_{n0}(x)/T_{n0}(x) \) is also the N-W estimator; for a boundary point \( x \) (i.e. \( x \in [0, h) \cup (1-h, 1] \)), \( s_1 \neq 0 \) and the estimator is different from the N-W estimator. Therefore, for \( p = 1 \) the plug-in estimator \( \hat{m}(x) \) can be regarded as boundary corrected N-W estimator, which shares some advantages with the N-W estimator, such as a simple closed formula for analysis and computation.

In the following, we focus only on the case with \( p = 1 \), which will show that the corresponding estimator \( \hat{m}(x) \) keeps its convergence rate for boundary points, i.e. the estimator automatically corrects the boundary effect.

4.2.1. Asymptotic Properties

The following theorem describes the asymptotic behavior and the boundary adaptation of the estimator \( \hat{m}(x) \) in (26).

**Theorem 4.1:** Suppose conditions (D1)-(D5) in Appendix I holds. Then for \( x \in [0, 1] \), \( \sqrt{nh}[\hat{m}(x) - m(x) - B(x)] \) is asymptotically normal with mean zero and variance

\[
D(x) = \frac{\sigma^2 \int_{d} (s_2 - us_1)^2 K^2(u) du}{f_1(x)(s_0 s_2 - s_1^2)^2},
\]

where

\[
B(x) = \frac{h^2 s_2^2 - s_1 s_3}{2 s_0 s_2 - s_1^2} \left[ m''(x) + 2 \frac{f'(x)m'(x)}{f_1(x)} \right] + o_p(h^2 + \frac{h}{\sqrt{nh}}).
\]

**Remark 4.1:** A consequence of this theorem is that the estimator \( \hat{m}(x) \) keeps its convergence rate for the boundary points. That is, the estimator is automatically boundary adaptive.

From Theorem 4.1, the asymptotic mean squared error (AMSE) for estimating \( m(x) \) is

\[
AMSE(h, x) = \frac{h^4(s_2^2 - s_1 s_3)^2}{4(s_0 s_2 - s_1^2)^2} [m''(x) + \frac{2m'(x)f'_1(x)}{f_1(x)}]^2
\]

\[
+ \frac{\sigma^2 \int_{d} (s_2 - us_1)^2 K^2(u) du}{nhf_1(x)(s_0 s_2 - s_1^2)^2}.
\]

(27)

Suppose \( x = ch, 0 < c < 1 \). Then for \( h \) small, \( d = 1 \), and (27) coincides with the AMSE expression for the locally linear estimator (18) except for the \( 2m'(x)f'_1(x)/f_1'(x) \) term (see Fan and Gijbels, p.17). Thus (26) is for
some models more efficient than the locally linear estimate at boundary and interior points. The optimal local bandwidth for estimating $m(x)$ at $x$, in the sense of minimizing $AMSE(h, x)$, is

$$h_{opt}(x) = n^{-\frac{1}{5}} \left( \frac{\sigma^2 \int_{c}^{d} (s_2 - u s_1)^2 K^2(u) du}{f_1(x) (s_2^2 - s_1 s_3)^2 [m''(x) + \frac{2m'(x) f_1'(x)}{f_1(x)}]^2} \right)^{\frac{1}{5}}.$$

**Acknowledgments**

This work was supported in part by the Chinese NSF grants 10001004 and 39930160, and by NSF grant DMS-9971301. We would like to thank Peter Bickel for valuable remarks on the early drafts of the paper.

**Appendix I: Conditions**

(A1) The density function $f(x)$ has a continuous $(p + 2)$th derivative at the point $x$ in its compact support $[0, 1]$, say.

(A2) The kernel function $K$ is a continuous function with bounded support $[-1, 1]$, say.

(A3) The weight function $w(\cdot)$ is nonzero and has a continuous derivative at the point $x$.

(A4) $h \to 0$ and $nh \to +\infty$ as $n \to +\infty$.

(B1) The density function $f(x)$ has a continuous second derivative at the point $x$ in its compact support $\Omega$.

(B2) The kernel function $K$ is a continuous function with bounded support.

(B3) $h \to 0$ and $nh \to +\infty$ as $n \to +\infty$.

(C1) The hazard function $\lambda(x)$ has a continuous $(p + 2)$th derivative at the point $x$.

(C2) The bandwidth $b$ tends to zero such that $nb \to +\infty$ as $n \to +\infty$. Let $B = \text{diag}(1, b, \ldots, b^p)$.

(C3) $L(x)$ is continuous at the point $x$.

(C4) The kernel function $K$ is a continuous function of bounded variation and with bounded support $[-1, 1]$.

(D1) The kernel function $K(x)$ is continuous with bounded support.

(D2) The density function $f(x)$ is bounded away from 0 and has a second derivative and compact support $[0, 1]$, say.
The second derivative of regression function \( m(x) \) is bounded and continuous.

The bandwidth sequence \( h_n \to 0 \) and satisfies \( nh_n \to \infty \) as \( n \to \infty \).

\[ \text{Var}(\varepsilon) = \sigma^2 < \infty, \text{ where } \varepsilon = Y - m(X). \]

Appendix II: Proof of the theorems

Proof of Theorem 2.1. By (3) and (4), we have

\[
B[\hat{\beta}(x) - \beta(x)] = C^{-1}[a(\hat{F}_n) - a(F)].
\]

Note that because \( \text{supp}(K) = [-1, 1] \), we need only consider \( |z - x| \leq b \) and \( |X_i - x| \leq b \). By a change of variable and the continuity of \( w(x) \)

\[
c_j = \int \left( \frac{z - x}{b} \right)^{j+\ell} K_b(z - x)w(z)dz = s_{j+\ell}w(x)(1 + o(1)),
\]

\[
a_\ell(F) = \int_{-c}^{d} t^\ell K(t) f(x + tb)w(x + tb)dt
\]

\[
= \int_{-c}^{d} \sum_{k=0}^{b^p+2} \frac{1}{k!} f^{(k)}(x)(tb)^k + o(b^{p+2})t^\ell K(t)w(x + tb)dt
\]

\[
= (c_0, \ell, \ldots, c_p, \ell)Bf + \sum_{j=1}^{2} b^{p+j}s_{\ell+p+j}w(x)f^{(p+j)}(x)/(p+j)! + o(b^{p+2}).
\]

Then \( C = Sw(x)(1 + o(1)) \), and

\[
a(F) = CBf + \sum_{j=1}^{2} b^{p+j}c_{p+j-1}w(x)f^{(p+j)}(x)/(p+j)! + o(b^{p+2}),
\]

so that

\[
B\beta(x) = C^{-1}a(F)
= BF + b^{p+1}S^{-1}c_pf(x)f^{(p+1)}(x)/(p+1)!
+ b^{p+2}S^{-1}c_{p+1}w(x)f^{(p+2)}(x)/(p+2)! + o(b^{p+2}). \quad (A.1)
\]

Note that

\[
a_\ell(\hat{F}) = n^{-1} \sum_{i=1}^{n} \left( \frac{X_i - x}{b} \right)^{\ell} K_b(X_i - x)w(X_i)
\]

is an i.i.d.’s sum. Applying the central limit theorem and the Cramér-Wold device, we obtain

\[
\sqrt{n}b[a(\hat{F}) - a(F)] \to N(0, V^*w^2(x)f(x)).
\]
Then \(\sqrt{nb}C^{-1}[a(\hat{F}) - a(F)] \to \mathcal{N}(0, S^{-1}V^*S^{-1}f(x))\), which combined with (A.1) completes the proof of the theorem. \(\diamondsuit\)

**Proof of Theorem 2.4.** Using the first order Taylor expansion and the same argument as for Theorem 2.1, we obtain the bias term

\[E\left[\hat{\alpha}(x) - f(x) \mid H[\hat{\beta}(x) - \nabla f(x)]\right] = \frac{1}{2}N_{x}^{-1}C_{x}(1 + o(1)). \tag{A.2}\]

It is straightforward to show that

\[\text{Cov}(A^{-1}S_n(x), A^{-1}S_n(x)) = n^{-1}|H|^{-1}f(x)T_x(1 + o(1)).\]

Note that \(A^{-1}S_n(x)\) is an i.i.d.’s sum, then using the central limit theorem and the Cramér-Wold device, one gets

\[\sqrt{n}|H|\left[\hat{\alpha}(x) - f(x) \mid H[\hat{\beta}(x) - \nabla f(x)]\right] \to \mathcal{N}(0, f(x)T_x).\]

This together with (A.2) yields the result of the theorem. \(\diamondsuit\)

**Proof of Corollary 2.1.** For an interior point \(x\), simple algebra gives

\[e^T_1N_{x}^{-1}C_{x} = \int K(u)u^TH\mathcal{H}_fHu\,du = \mu_2(K)tr(H\mathcal{H}_fH),\]

and \(e^T_1N_{x}^{-1}T_xN_{x}^{-1}e_1 = \int K^2(u)\,du\). The corollary now follows from Theorem 2.4. \(\diamondsuit\)

**Proof of Theorem 3.2.** Note that (14) is for estimating the density function \(\ell(t)\) for the uncensored observations, by Theorem 2.1 we have

\[\sqrt{nb}\{[\ell_1(t) - \ell_1(t)] - \frac{\ell^{(p+1)}(t)b^{p+1}}{(p + 1)!}S^{-1}_c\varepsilon + o(b^{p+1})\}\]

\[\to \mathcal{N}(0, e^T_1S^{-1}V^*S^{-1}e_1\ell_1(t)).\]

This combined with \(\hat{L}(t) = \tilde{L}(t) + O_p(1/\sqrt{n})\) yields the result of the theorem. \(\diamondsuit\)

**Proof of Theorem 4.1.** Note that

\[S_{nt}(x) = n^{-1} \sum_{i=1}^{n} (X_i - x)^\ell h^{-\ell}K_h(X_i - x)Y_i \]

\[= n^{-1} \sum_{i=1}^{n} (X_i - x)^\ell h^{-\ell}K_h(X_i - x) [m(X_i) + \varepsilon_i] \equiv B_{nt} + V_{nt}.\]
By Taylor expansion, one gets
\[
\hat{m}(x) - m(x) = \frac{s_2(B_{n0} - m(x)T_{n0}) - s_1(B_{n1} - m(x)T_{n1})}{s_2T_{n0} - s_1T_{n1}} + \frac{s_2V_{n0} - s_1V_{n1}}{s_2T_{n0} - s_1T_{n1}}
\equiv B_n + V_n.
\]

We will show that the first term above contributes the bias of the estimator \( \hat{m}(x) \) and the second term the variance:

Since \( K(\cdot) \) vanishes outside of \([-1, 1]\), we need only consider \( |X_i - x| \leq h \). Note that for \( \ell = 0, 1 \)
\[
B_{n\ell} - m(x)T_{n\ell} = n^{-1} \sum_{i=1}^{n} (X_i - x)^\ell h^{-\ell} K_h(X_i - x) \left[ m(X_i) - m(x) \right].
\]  
(A.3)

By Taylor expansion, one gets
\[
m(X_i) - m(x) = \sum_{k=1}^{2} h^{k} m^{(k)}(x)/k! \left[ (X_i - x)/h \right]^{k} + o(h^{2}),
\]
uniformly for \( |X_i - x| \leq h \). This combined with (A.3) yields
\[
B_{n\ell} - m(x)T_{n\ell} = h m'(x)n^{-1} \sum_{i=1}^{n} [(X_i - x)/h]^{\ell+1} K_h(X_i - x) + \frac{h^{2}m''(x)}{2}n^{-1} \sum_{i=1}^{n} [(X_i - x)/h]^{\ell+2} K_h(X_i - x) + o(h^{\ell+2}).
\]

By directly computing the mean and variance we have
\[
B_{n\ell} - m(x)T_{n\ell} = h s_{\ell+1}m'(x)f(x) + h^{2} s_{\ell+2}[m'(x)f'(x) + \frac{1}{2}m''(x)f(x)] + o(h^{2}) + O_p(h/\sqrt{nh}).
\]  
(A.4)

Note that
\[
s_2T_{n0} - s_1T_{n1} = (s_2s_0 - s_1^2)f(x) + \frac{h^2}{2}(s_2^2 - s_1s_3)f''(x)
+ o(h^2) + O_p(h/\sqrt{nh}).
\]
(A.5)

It follows from (A.4) and (A.5) that
\[
B_n = \frac{h^2 s_2^2 - s_1 s_3}{2 s_0 s_2 - s_1^2} m''(x)f(x) + 2m'(x)f'(x) + o(h^2) + O_p(h/\sqrt{nh}).
\]
Plug-in Estimation

Note that $E(V_{n\ell}) = 0$ and by a change of variable
\[
\text{Var}(V_{n\ell}) = \frac{\sigma^2}{nh^2}E[(X_i - x)^2h^{-2}\sum_s K^2(\frac{t-s}{h})]
\]
\[
= \frac{\sigma^2}{nh^2} \int_0^1 (t - x)^2h^{-2}\sum_s K^2(\frac{t-x}{h})f(t)dt
\]
\[
= \frac{\sigma^2}{nh} f(x) \int_{-c}^d u^2 K^2(u)du(1 + o(1)). \tag{A.6}
\]

and
\[
\text{Cov}(V_{n0}, V_{n1}) = \frac{\sigma^2}{nh^2} \int_0^1 \frac{t-x}{h}K^2(\frac{t-x}{h})f(t)dt
\]
\[
= \frac{\sigma^2}{nh} f(x) \int_{-c}^d uK^2(u)du(1 + o(1)). \tag{A.7}
\]

It follows from (A.6) and (A.7) that
\[
\text{Var}(s_2V_{n0} - s_1V_{n1}) = \frac{\sigma^2}{nh} f(x) \int_{-c}^d (s_2 - us_1)^2 K^2(u)du(1 + o(1)). \tag{A.8}
\]

Since $s_2V_{n0} - s_1V_{n1}$ is a sum of iid random variables of mean zero, it is asymptotically normal with mean zero and variance as in (A.8). This together with (A.5) gives the variance of $V_n$. ⊁

References