Convex Surrogate Minimization in Classification

Cui Xiong[1], Jun Shao[1,2] and Lei Wang[3]

[1]*East China Normal University,* [2]*University of Wisconsin-Madison*

and [3]*Nankai University*

Supplementary Material

Proof of Lemma 1. For j being an integer with $2 \le j \le p$, let β_j and β_{0j} be the jth components of β and β_0, respectively. Note that the element β_{0p} is assumed to be non-zero. Assume it is positive. Taking the derivative of $R(\beta)$ with respect to β_j, we have

$$\begin{aligned}
&\frac{\partial}{\partial\beta_j}E\{\ell(Y\beta^TX)\}\\
&=\frac{\partial}{\partial\beta_j}\int_{x_j\le-\frac{(\beta^Tx)_{-j}}{\beta_j}}g(\beta_0^Tx)f(x)dx+\frac{\partial}{\partial\beta_j}\int_{x_j\ge-\frac{(\beta^Tx)_{-j}}{\beta_j}}\{1-g(\beta_0^Tx)\}f(x)dx\\
&=\int g\left((\beta_0^Tx)_{-j}-\frac{(\beta^Tx)_{-j}\beta_{0,j}}{\beta_j}\right)f\left(x_{-j},-\frac{(\beta^Tx)_{-j}}{\beta_j}\right)\frac{(\beta^Tx)_{-j}}{\beta_j^2}dx_{-j}\\
&+\int\left[1-g\left((\beta_0^Tx)_{-j}-\frac{(\beta^Tx)_{-j}\beta_{0,j}}{\beta_j}\right)\right]f\left(x_{-j},-\frac{(\beta^Tx)_{-j}}{\beta_j}\right)\frac{-(\beta^Tx)_{-j}}{\beta_j^2}dx_{-j}
\end{aligned}$$

where $x_{-j}=(x_2,...,x_{j-1},x_{j+1},...,x_p)$ and $(\beta^Tx)_{-j}=\beta_1+\beta_2x_2+\cdots+\beta_{j-1}x_{j-1}+\beta_{j+1}x_{j+1}+\cdots+\beta_px_p$. The first equation follows from the fact that β appears in the limits of integrals only. Then,

$$\begin{aligned}\frac{\partial}{\partial\beta_j}E\{\ell(Y\beta^TX)\}\Big|_{\beta=c\beta_0} &= \int g(0)f\left(x_{-j}, -\frac{(\beta_0^Tx)_{-j}}{\beta_{0j}}\right)\frac{(\beta_0^Tx)_{-j}}{c\beta_{0j}^2}dx_{-j}\\ &+\int\{1-g(0)\}f\left(x_{-j}, -\frac{(\beta_0^Tx)_{-j}}{\beta_{0j}}\right)\frac{-(\beta_0^Tx)_{-j}}{c\beta_{0j}^2}dx_{-j}\end{aligned}$$

which is 0 because $g(0) = 1 - g(0)$. Similarly, for the intercept β_1,

$$\begin{aligned}&\frac{\partial}{\partial\beta_1}E\{\ell(Y\beta^TX)\}\\ &= \frac{\partial}{\partial\beta_1}\int_{x_p\le -\frac{(\beta^Tx)_{-p}}{\beta_p}} g(\beta_0^Tx)f(x)dx + \frac{\partial}{\partial\beta_1}\int_{x_p\ge -\frac{(\beta^Tx)_{-p}}{\beta_p}}\{1-g(\beta_0^Tx)\}f(x)dx\\ &= \int g\left((\beta_0^Tx)_{-p} - \frac{(\beta^Tx)_{-p}\beta_{0,p}}{\beta_p}\right)f\left(x_{-p}, -\frac{(\beta^Tx)_{-p}}{\beta_p}\right)\frac{1}{\beta_p}dx_{-p}\\ &\quad -\int\left[1-g\left((\beta_0^Tx)_{-p} - \frac{(\beta^Tx)_{-p}\beta_{0,p}}{\beta_p}\right)\right]f\left(x_{-p}, -\frac{(\beta^Tx)_{-p}}{\beta_p}\right)\frac{1}{\beta_p}dx_{-p}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial\beta_1}E\{\ell(Y\beta^TX)\}\Big|_{\beta=c\beta_0} &= \int g(0)f\left(x_{-p}, -\frac{(\beta_0^Tx)_{-p}}{\beta_{0p}}\right)\frac{1}{c\beta_{0p}}dx_{-p}\\ &-\int\{1-g(0)\}f\left(x_{-p}, -\frac{(\beta_0^Tx)_{-p}}{\beta_{0p}}\right)\frac{1}{c\beta_{0p}}dx_{-p},\end{aligned}$$

which is 0 because $g(0) = 1 - g(0)$. This proves Lemma 1. $\square$

Proof of Lemma 2. Note that

$$\begin{aligned}R_\varphi(\beta) = E\{\varphi(Y\beta^TX)\} &= E\left[E\{\varphi(Y\beta^TX)|X\}\right]\\ &= E\left[g(\beta_0^TX)\varphi(\beta^TX) + \{1-g(\beta_0^TX)\}\varphi(-\beta^TX)\right]\\ &= \int\left[\varphi(\beta^Tx)g(\beta_0^Tx) + \varphi(-\beta^Tx)\{1-g(\beta_0^Tx)\}\right]dF(x)\end{aligned}$$

and, hence, by the Dominated Convergence Theorem, we have

$$\frac{\partial R_\varphi(\beta)}{\partial\beta} = \int \left[\varphi'(\beta^T x)g(\beta_0^T x) - \varphi'(-\beta^T x)\{1-g(\beta_0^T x)\}\right] x dF(x).$$

If $\varphi \in \Psi(\beta_0)$, by (9), $\frac{\partial R_\varphi(\beta)}{\partial\beta}|_{\beta=\beta_0} = 0$. Since

$$\frac{\partial^2 R_\varphi(\beta)}{\partial\beta\partial\beta^T} = \int \left[\varphi''(\beta^T x)g(\beta_0^T x) + \varphi''(-\beta^T x)\{1-g(\beta_0^T x)\}\right] xx^T dF(x),$$

$\frac{\partial^2 R_\varphi(\beta)}{\partial\beta\partial\beta^T}|_{\beta=\beta_0}$ is positive definite. Hence, β_0 is the unique minimizer of $R_\varphi(\beta)$.

□

Proof of Theorem 2. (i) Define

$$L(\beta) = \frac{1}{n}\sum_{i=1}^n \tilde\varphi'(Y_i\beta^T X_i)Y_i X_i K_h(\beta^T X_i).$$

We first show that

$$E\{L(g'(0)\beta_0)\} \asymp h^2. \tag{S0.1}$$

Consider the surrogate $\tilde\varphi$ in (5) and let $U = \{g'(0)\beta_0^T X\}/h$. Then

$$\begin{aligned}
&\frac{1}{2}E\left[\tilde\varphi'(Yg'(0)\beta_0^T X)YXK_h(g'(0)\beta_0^T X)\right]\\
&= E\left[\left\{Yg'(0)\beta_0^T X - \frac{1}{2}\right\}YXK_h(g'(0)\beta_0^T X)\right]\\
&= E\left[\left\{g'(0)\beta_0^T X - \frac{1}{2}(2g(\beta_0^T X)-1)\right\}XK_h(g'(0)\beta_0^T X)\right]\\
&= E\left[\left\{hU + \frac{1}{2} - g\left(\frac{hU}{g'(0)}\right)\right\}XK(U)/h\right]\\
&= E\left[\left\{hU + \frac{1}{2} - g(0) - g'(0)\frac{hU}{g'(0)} + g''(0)\frac{h^2U^2}{2g'^2(0)} - g'''(\xi)\frac{h^3U^3}{6g'^3(0)}\right\}XK(U)/h\right]\\
&= \frac{h^2 g''(0)}{2g'^2(0)}E\left[U^2X\frac{K(U)}{h}\right] - \frac{h^3}{6g'^3(0)}E\left[g'''(\xi)U^3X\frac{K(U)}{h}\right],
\end{aligned} \tag{S0.2}$$

where ξ is between 0 and $hU/g'(0)$. Consider the transformation

$$u = g'(0)\frac{\beta_{01} + \beta_{02}x_2 + \cdots + \beta_{0p}x_p}{h}, \quad \text{and} \quad x_j = x_j,\ j = 2, ..., p-1.$$

Let $dx_{-p} = dx_2 \cdots dx_{p-1}$. For $j = 2, ..., p-1$, the jth component of $E\left[U^2 X \frac{K(U)}{h}\right]$ is the integral

$$\begin{aligned}
&\frac{1}{h}\int_{u\in[-1,1]} u^2 x_j K(u) f(x_2, ..., x_p) dx_2 \cdots dx_p \\
&= \int_{u\in[-1,1]} u^2 x_j K(u) f\left(x_2, ..., x_{p-1}, \frac{uh}{\beta_{0p}g'(0)} - \frac{(\beta_0^T x)_{-p}}{\beta_{0p}}\right) \frac{1}{|\beta_{0p}|g'(0)} du dx_{-p} \\
&\xrightarrow{h\to 0} \int_{u\in[-1,1]} u^2 x_j K(u) f\left(x_2, ..., x_{p-1}, -\frac{(\beta_0^T x)_{-p}}{\beta_{0p}}\right) \frac{1}{|\beta_{0p}|g'(0)} du dx_{-p} \\
&= \frac{B_k}{|\beta_{0p}|g'(0)} \int x_j f\left(x_2, ..., x_{p-1}, -\frac{(\beta_0^T x)_{-p}}{\beta_{0p}}\right) dx_{-p} \\
&= \frac{B_k}{|\beta_{0p}|g'(0)} \int x_j f\left(z\right) dx_{-p},
\end{aligned}$$

where $z = (x_2, ..., x_{p-1}, -(\beta_0^T x)_{-p}/\beta_{0p})^T$. Similarly, the first component of $E\left[U^2 X \frac{K(U)}{h}\right]$ is $\frac{B_k}{|\beta_{0p}|g'(0)} \int f\left(z\right) dx_{-p}$ and the pth component of $E\left[U^2 X \frac{K(U)}{h}\right]$ is the integral

$$\begin{aligned}
&\frac{1}{h}\int_{u\in[-1,1]} u^2 x_p K(u) f(x_2, ..., x_p) dx_2 \cdots dx_p \\
&= \frac{1}{|\beta_{0p}|g'(0)} \int_{u\in[-1,1]} u^2 K(u) \frac{uh - (\beta_0^T x)_{-p} g'(0)}{\beta_{0p}g'(0)} \\
&\qquad\qquad \times f\left(x_2, ..., x_{p-1}, \frac{uh - (\beta_0^T x)_{-p} g'(0)}{\beta_{0p}g'(0)}\right) du dx_{-p} \\
&\to \frac{1}{|\beta_{0p}|g'(0)} \int_{u\in[-1,1]} u^2 \left(-\frac{(\beta_0^T x)_{-p}}{\beta_{0p}}\right) K(u) f\left(x_2, ..., x_{p-1}, -\frac{(\beta_0^T x)_{-p}}{\beta_{0p}}\right) du dx_{-p} \\
&= \frac{1}{|\beta_{0p}|g'(0)} B_k \int \left(-\frac{(\beta_0^T x)_{-p}}{\beta_{0p}}\right) f\left(z\right) dx_{-p}.
\end{aligned}$$

Combining these results, we obtain that each component of the first term on the right hand side of (S0.2) $\asymp h^2$. The jth component of the second term on the right hand side of (S0.2) is bounded by

$$\frac{h^3 \max_x |g'''(x)|}{6g'^3(0)} E \left| U^3 X \frac{K(U)}{h} \right|$$

Replacing $u^2 x_j$ by $|u^3 x_j|$ in the previous proof we obtain that each component of the second term on the right hand side of (S0.2) $\asymp h^3$. Hence,

$$\text{each component of } E\left[\left\{Yg'(0)\beta_0^T X - \frac{1}{2}\right\} YXK_h(g'(0)\beta_0^T X)\right] \asymp h^2.$$

To prove (i), we also need to calculate $\frac{\partial L(\beta)}{\partial \beta}\big|_{\beta=g'(0)\beta_0}$ and find the asymptotic distribution of $L(g'(0)\beta_0)$. Note that

$$\begin{aligned}
\frac{1}{2}\frac{\partial L(\beta)}{\partial \beta}\bigg|_{\beta=g'(0)\beta_0} &= \frac{1}{n}\sum_{i=1}^n X_i X_i^T K_h(g'(0)\beta_0^T X_i) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \left(Y_i g'(0)\beta_0^T X_i - \frac{1}{2}\right) Y_i X_i X_i^T K_h'(g'(0)\beta_0^T X_i).
\end{aligned}$$

Using almost the same proof as that for (S0.1), we obtain that

$$\begin{aligned}
& E\left\{\left(Yg'(0)\beta_0^T X - \frac{1}{2}\right) YXX^T K_h'(g'(0)\beta_0^T X)\right\} \\
&= E\left\{\left[g'(0)\beta_0^T X - \frac{1}{2}\{2g(\beta_0^T X) - 1\}\right] XX^T K_h'(g'(0)\beta_0^T X)\right\} \\
&= E\left\{\left[hU + \frac{1}{2} - g\left(\frac{hU}{g'(0)}\right)\right] XX^T K'(U)/h\right\} \\
&= E\left[\left\{g''(0)\frac{h^2U^2}{2g'^2(0)} - g'''(\xi)\frac{h^3U^3}{6g'^3(0)}\right\} XX^T K'(U)/h\right] \\
&\to 0.
\end{aligned}$$

On the other hand,

$$
\begin{aligned}
&E\{XX^T K_h(g'(0)\beta_0^T X)\} = \frac{1}{h} E\{XX^T K(U)\} \\
&= \frac{1}{h} \int_{u\in[-1,1]} xx^T K(u) f(x_2, ..., x_p) dx_2 \cdots dx_p \\
&\to \frac{1}{|\beta_{0p}| g'(0)} \int_{u\in[-1,1]} \begin{pmatrix} 1 & z^T \\ z & zz^T \end{pmatrix} K(u) f\left(x_2, ..., x_{p-1}, -\frac{(\beta_0^T x)_{-p}}{\beta_{0p}}\right) du dx_{-p} \\
&= D
\end{aligned}
$$

for the D defined in (16). By the law of large numbers,

$$
\left.\frac{\partial L(\beta)}{\partial \beta}\right|_{\beta = g'(0)\beta_0} \to 2D
$$

in probability. We further calculate the covariance matrix of $L(g'(0)\beta_0)$. From (S0.1), we have $E\{L(g'(0)\beta_0)\} \asymp h^2$. Then

$$
\begin{aligned}
&\mathrm{Cov}\{L(g'(0)\beta_0)\} \\
&= \frac{4}{nh^2} E\left[\left\{Y g'(0)\beta_0^T X - \frac{1}{2}\right\}^2 XX^T K^2\left(\frac{g'(0)\beta_0^T X}{h}\right)\right] \\
&\quad - \frac{1}{n} E\{L(g'(0)\beta_0) L(g'(0)\beta_0)^T\} \\
&= \frac{4}{nh^2} E\left[\left\{\frac{1}{4} - \left(2g(\beta_0^T X) - 1\right) g'(0)\beta_0^T X + \left(g'(0)\beta_0^T X\right)^2\right\} XX^T K^2\left(\frac{g'(0)\beta_0^T X}{h}\right)\right] \\
&\quad - \frac{1}{n} E\{L(g'(0)\beta_0) L(g'(0)\beta_0)^T\} \\
&= \frac{4}{nh^2} E\left[\left\{\frac{1}{4} - h^2 U^2 - g''(0)\frac{h^3 U^3}{g'^2(0)} - g'''(\xi)\frac{h^4 U^4}{3g'^3(0)}\right\} XX^T K^2(U)\right] \\
&\quad - \frac{1}{n} E\{L(g'(0)\beta_0) L(g'(0)\beta_0)^T\}.
\end{aligned}
$$

Using the same argument as before, we obtain that

$$\begin{aligned}
&\frac{1}{h}E\left[\left\{\frac{1}{4}-h^2U^2-g''(0)\frac{h^3U^3}{g'^2(0)}-g'''(\xi)\frac{h^4U^4}{3g'^3(0)}\right\}XX^TK^2(U)\right]\\
&=\frac{1}{h}\int_{u\in[-1,1]}\left\{\frac{1}{4}-h^2U^2-g''(0)\frac{h^3U^3}{g'^2(0)}-g'''(\xi)\frac{h^4U^4}{3g'^3(0)}\right\}xx^T\\
&\qquad\qquad\qquad\qquad\times K^2(u)f(x_2,...,x_p)dx_2\cdots dx_p\\
&\to\int_{u\in[-1,1]}\frac{K^2(u)}{4|\beta_{0p}|g'(0)}\begin{pmatrix}1 & z^T\\ z & zz^T\end{pmatrix}f\left(x_2,...,x_{p-1},-\frac{(\beta_0^Tx)_{-p}}{\beta_{0p}}\right)dudx_{-p}\\
&=\frac{V_kD}{4}.
\end{aligned}$$

This shows that

$$\mathrm{Cov}\{L(g'(0)\beta_0)\}\asymp 1/(nh),$$

since $E\{L(g'(0)\beta_0)L(g'(0)\beta_0)^T\}\asymp h^4$. By the central limit theorem,

$$\sqrt{nh}[L(g'(0)\beta_0)-E\{L(g'(0)\beta_0)\}]\to N_p(0,V_kD)$$

in distribution. Since $E\{L(g'(0)\beta_0)\}\asymp h^2$, we have

$$\sqrt{nh}L(g'(0)\beta_0)\to N_p(0,V_kD) \tag{S0.3}$$

in distribution, under the assumed condition that $nh^5\to 0$.

Based on the minimum distance theory (Newey and Mcfadden, 1994), let $Q(\beta)=L(\beta)^TL(\beta)$ and define

$$\hat\beta=\mathrm{argmin}_\beta Q(\beta).$$

The local identification can be verified since $\frac{\partial L(\beta)}{\partial \beta}\big|_{\beta=g'(0)\beta_0}$ is positive definite. Next, to show $\hat{\beta} \to g'(0)\beta$ in probability, the proof is similar to that of Lemma 1 of Qin and Lawless (1994). Denote $\beta = g'(0)\beta_0 + u(nh)^{-1/3}$ for $\beta \in \{\beta \mid \|\beta - g'(0)\beta_0\| = (nh)^{-1/3}\}$, where $\|u\| = 1$ and $\|\cdot\|$ denotes Euclidean norm.

First, we give a lower bound for $Q(\beta)$ when β belongs to the ball $\|\beta - g'(0)\beta_0\| \leq (nh)^{-1/3}$. By Taylor expansion and (S0.3), we have (uniformly for u)

$$\begin{aligned}
Q(\beta) &= \{L(g'(0)\beta_0) + L'(g'(0)\beta_0)u(nh)^{-1/3}\}^T\{L(g'(0)\beta_0) \\
&\qquad + L'(g'(0)\beta_0)u(nh)^{-1/3}\} + o((nh)^{-2/3}) \\
&= \{O((nh)^{-1/2}) + L'(g'(0)\beta_0)u(nh)^{-1/3}\}^T\{O((nh)^{-1/2}) \\
&\qquad + L'(g'(0)\beta_0)u(nh)^{-1/3}\} + o((nh)^{-2/3}) \\
&\geq C \cdot (nh)^{-2/3},
\end{aligned}$$

with $C > 0$. Similarly, we have $Q(g'(0)\beta_0) = O((nh)^{-1})$. Since $Q(\beta)$ is a continuous function about β as β belongs to the ball $\|\beta - g'(0)\beta_0\| \leq (nh)^{-1/3}$, with probability tending to 1, $Q(\beta)$ has a minimum $\hat{\beta}$ in the interior of the ball, and this $\hat{\beta}$ satisfies

$$\frac{\partial Q(\beta)}{\partial \beta}\Big|_{\beta=\hat{\beta}} = 2\frac{\partial L(\beta)^T}{\partial \beta}\Big|_{\beta=\hat{\beta}} L(\hat{\beta}) = 0,$$

which holds only when $L(\hat{\beta}) = 0$. That is with probability tending to 1,

$L(\beta) = 0$ has a root in the interior of the ball $\|\beta - g'(0)\beta_0\| \leq (nh)^{-1/3}$.

To prove (ii), by Taylors expansion, there exists a η between $\hat{\beta}$ and $g'(0)\beta_0$ such that

$$L(\hat{\beta}) - L(g'(0)\beta_0) = \frac{\partial L(\beta)}{\partial \beta}\Big|_{\beta=\eta}(\hat{\beta} - g'(0)\beta_0),$$

which implies that

$$(nh)^{1/2}\{\hat{\beta} - g'(0)\beta_0\} = -(nh)^{1/2}\Big\{\frac{\partial L(\beta)}{\partial \beta}\Big|_{\beta=\eta}\Big\}^{-1} L(g'(0)\beta_0).$$

Using the fact that $\hat{\beta} \to g'(0)\beta_0$ in probability, we have $\big\{\frac{\partial L(\beta)}{\partial \beta}\big|_{\beta=\eta}\big\}^{-1} \to (2D)^{-1}$ in probability, which also implies that

$$(nh)^{1/2}\{\hat{\beta} - g'(0)\beta_0\} = -(nh)^{1/2}(2D)^{-1}L(g'(0)\beta_0).$$

That is the asymptotic distribution of $(nh)^{1/2}\{\hat{\beta} - g'(0)\beta_0\}$ is the same as the asymptotic distribution of $-(nh)^{1/2}(2D)^{-1}L(g'(0)\beta_0)$. Therefore,

$$\sqrt{nh}\{\hat{\beta} - g'(0)\beta_0\} \to N_p\left(0, V_k D^{-1}/4\right)$$

in distribution. This proves the results in (i)-(ii).

From the proofs of (i)-(ii), the bias of $\hat{\beta}$ as an estimator of $g'(0)\beta_0$ is of the order h^2 and the covariance matrix of $\hat{\beta}$ is of the order $(nh)^{-1}$. Hence, the asymptotic mean squared error of $\hat{\beta}$ is of the order $(nh)^{-1} + h^4$. Therefore, the best rate of convergence to 0 in mean squared error is achieved when $h \asymp n^{-1/5}$. This proves part (iii) of the theorem.

For $j = 1, ..., p-1$ and $k = 1, ..., p-1$,

$$
\begin{aligned}
&\frac{\partial^2 R(\beta)}{\partial\beta_j\partial\beta_k} \\
&= \frac{\partial^2}{\partial\beta_j\partial\beta_k}\int_{x_p\le -\frac{(\beta^T x)_{-p}}{\beta_p}} g(\beta_0^T x)f(x)dx + \frac{\partial^2}{\partial\beta_j\partial\beta_k}\int_{x_p\ge -\frac{(\beta^T x)_{-p}}{\beta_p}} \{1-g(\beta_0^T x)\}f(x)dx \\
&= \frac{\partial}{\partial\beta_j}\int \left[1-2g\left((\beta_0^T x)_{-p} - \frac{\beta_{0p}}{\beta_p}(\beta^T x)_{-p}\right)\right] f\left(x_{-p}, -\frac{1}{\beta_p}(\beta^T x)_{-p}\right)\frac{x_k}{\beta_p}dx_{-p} \\
&= \int \frac{2\beta_{0p}x_j x_k}{\beta_p^2} g'\left((\beta_0^T x)_{-p} - \frac{\beta_{0p}}{\beta_p}(\beta^T x)_{-p}\right) f\left(x_{-p}, -\frac{1}{\beta_p}(\beta^T x)_{-p}\right)dx_{-p} \\
&+ \int \left[1-2g\left((\beta_0^T x)_{-p} - \frac{\beta_{0p}}{\beta_p}(\beta^T x)_{-p}\right)\right]\frac{\partial}{\partial\beta_j}\left\{\frac{x_k}{\beta_p} f\left(x_{-p}, -\frac{1}{\beta_p}(\beta^T x)_{-p}\right)\right\}dx_{-p}
\end{aligned}
$$

It is clear that

$$\left.\frac{\partial^2 R(\beta)}{\partial\beta\partial\beta^T}\right|_{\beta=g'(0)\beta_0} = 2D$$

Since, $\hat{\beta} - g'(0)\beta_0 = O_p\left((nh)^{-1/2}\right)$ and

$$R(\hat{\beta}) - R(g'(0)\beta_0) = \frac{1}{2}\left(\hat{\beta} - g'(0)\beta_0\right)^T \left.\frac{\partial^2 R(\beta)}{\partial\beta\partial\beta^T}\right|_{\beta=g'(0)\beta_0}\left(\hat{\beta} - g'(0)\beta_0\right)\{1+o_p(1)\},$$

we have

$$R(\hat{\beta}) - R(g'(0)\beta_0) = O_p\left((nh)^{-1}\right).$$

This proves part (iv) of the theorem. □

References

Newey, W. and Mcfadden, D. (1994). *Large Sample Estimation and Hypothesis Testing*. Springer, New York.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* **22**, 300-325.

Cui Xiong

School of Statistics, East China Normal University,

Shanghai 200241, China.

E-mail:cxiong531@163.com

Jun Shao

Department of Statistics, University of Wisconsin-Madison,

1300 University Ave., Madison, Wisconsin, 53706, U.S.A.

E-mail: shao@stat.wisc.edu

Lei Wang

LPMC and Institute of Statistics, Nankai University,

Tianjin 300071, China.

E-mail:leiwang.stat@gmail.com