Appendix B  Proofs of Theorems 1 - 3. (PDF file)

Proof of Theorem 1.

Let us consider \( \theta_0 \) implicitly defined by (2) and fix some other \( \theta \). Then, using Jensen’s inequality, (5) and (10),

\[
I(\theta) - I(\theta_0) = E(I_n(\theta) - I_n(\theta_0)) = E\left( \log \left( \frac{f_\theta(Y|\theta'X)}{f_{\theta_0}(Y|\theta_0'X)} \right) \right) \\
\leq \log \left( E \left( \frac{f_\theta(Y|\theta'X)}{f_{\theta_0}(Y|\theta_0'X)} \right) \right) \\
= \log \left( \int \int \frac{f_\theta(y|\theta'x)}{f_{\theta_0}(y|\theta_0'x)} f_{\theta_0}(y|\theta_0'x)f_X(x)dx dy \right) \\
= \log \left( \int \left( \int f_\theta(y|\theta'x)dy \right) f_X(x)dx \right) = \log 1 = 0.
\]

So \( I(\theta) \leq I(\theta_0) \) \( \forall \theta \), which completes the proof. \( \square \)

Proof of Lemma 1. Using (11) and (9) we have

\[
l(\theta_0) = 0 \quad \text{and} \quad I_n(\hat{\theta}_n) = 0.
\]

Now a Taylor expansion gives

\[
l(\theta_0) = 0 = I_n(\hat{\theta}_n) = I_n(\hat{\theta}_n) + I_n^{(0)}(\hat{\theta}_n) - I_n^{(0)}(\hat{\theta}_n),
\]

where \( \hat{\theta}_n \) is between \( \theta_0 \) and \( \theta_0 \). This completes the proof. \( \square \)

The proof of Theorem 2 is postponed. Let us first state and prove an auxiliary lemma.

Lemma 2. Under the conditions in Theorem 3 we have

\[
i_n^{(0)}(\theta_0) = I_n^{(0)}(\theta_0) \to 0
\]

and

\[
\sqrt{n}(I_n^{(0)}(\theta_0) - I^{(0)}(\theta_0)) \to N(0, \Sigma_1)
\]

Proof. According to (5) and since \( I^{(0)}(\theta_0) = E(I_n^{(0)}(\theta_0)) = 0 \), we have

\[
\sqrt{n}(I_n^{(0)}(\theta_0) - I^{(0)}(\theta_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{f_\theta(Y_i|\theta'X_i)_{\theta=\theta_0}}{f_{\theta_0}(Y_i|\theta_0'X_i)} \right]
\]
Then, \( \tilde{l}_n^{[1]}(\theta) - l^{[1]}(\theta) \) is a sum of i.i.d. random vectors and the law of large numbers gives the convergence to zero in probability. Now the central limit theorem leads to

\[
\sqrt{n}(\tilde{l}_n^{[1]}(\theta_0) - l^{[1]}(\theta_0)) \rightarrow \mathcal{N}(0, \Sigma_1)
\]

The matrix \( \Sigma_1 \) is easily computed:

\[
\Sigma_1 = E \left[ \left( \frac{f^{[1]}_\theta(Y_1|\theta'X_1)}{f_\theta(Y_1|\theta'X_1)} \right) \left( \frac{f^{[1]}_\theta(Y_1|\theta'X_1)}{f_\theta(Y_1|\theta'X_1)} \right)' \right]
\]

\[
= \int (\nabla_\theta \log(f_\theta(z|\theta'x)))_{\theta = \theta_0} (\nabla_\theta \log(f_\theta(z|\theta'x)))_{\theta = \theta_0}' f(x, z) dx dz.
\]

Observe that the last equation is a consequence of Lemma 4 below. \( \square \)

**Proof of Theorem 3.** In view of Lemma 1, the term

\[
\tilde{l}_n^{[1]}(\theta_0) - l^{[1]}(\theta_0) = \alpha_n(\theta_0) + \beta_n(\theta_0),
\]

has to be studied, where

\[
(20) \quad \alpha_n(\theta_0) = \tilde{l}_n^{[1]}(\theta_0) - \tilde{l}_n^{[1]}(\theta_0)
\]

\[
\beta_n(\theta_0) = l^{[1]}(\theta_0) - l^{[1]}(\theta_0).
\]

To deal with \( \beta_n(\theta_0) \), we use Lemma 2, to prove

\[
\sqrt{n} \beta_n(\theta_0) \rightarrow \mathcal{N}(0, \Sigma_1)
\]

and

\[
\beta_n(\theta_0) \rightarrow 0 \quad \text{in probability.}
\]

Concerning \( \alpha_n(\theta) \), using (5), (8) and (6), we have

\[
\tilde{l}_n(\theta_0) - \tilde{l}_n(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} \left[ (\log(\tilde{r}^{-1}(\theta_0'X_i, Y_i)) - \log(f_{\theta_0}(\theta_0'X_i, Y_i))) + (\log(f_{\theta_0}(\theta_0'X_i)) - \log(\tilde{s}^{-1}(\theta_0'X_i))) \right],
\]
which, in view of (20), implies

\[ \hat{\eta}^{[1]}(\theta_0) - \hat{\eta}^{[1]}(\theta_0) = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \hat{\tau}^{-i}[\theta_0](x_i, y_i) - \frac{f_0^{[1]}(x_i, y_i)}{f_0(x_i)} \right) + \frac{f_0^{[1]}(x_i)}{f_0(x_i)} \frac{\hat{s}^{-i}[\theta_0](x_i)}{\hat{s}(x_i)} \right] \]

Now using

\[ \frac{1}{\hat{\tau}^{-i}} = \frac{1}{f_0} + \frac{f_0 - \hat{\tau}^{-i}}{f_0 \hat{\tau}^{-i}}, \quad \frac{1}{\hat{s}^{-i}} = \frac{1}{f_0^{[0]}(x_i)} + \frac{f_0^{[0]}(x_i) - \hat{s}^{-i}}{f_0^{[0]}(x_i) \hat{s}^{-i}} \]

we obtain

\[ \hat{\eta}^{[1]}(\theta_0) - \hat{\eta}^{[1]}(\theta_0) = \sum_{k=1}^{8} A_{kn} + o_P(n^{-1/2}), \]

where

\[
\begin{align*}
A_{1n} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\tau}^{-i}[\theta_0](x_i, y_i) - f_0^{[1]}(x_i, y_i)}{f_0(x_i)} \\
A_{2n} &= \frac{1}{n} \sum_{i=1}^{n} \frac{f_0^{[1]}(x_i, y_i)}{f_0(x_i)} (f_0(x_i) - \hat{\tau}^{-i}(x_i, y_i)) \\
A_{3n} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{s}^{-i}[\theta_0](x_i)}{f_0^{[0]}(x_i)} \\
A_{4n} &= \frac{1}{n} \sum_{i=1}^{n} \frac{f_0^{[1]}(x_i)}{f_0^{[0]}(x_i)} (\hat{s}^{-i}(x_i, y_i) - f_0^{[0]}(x_i)) \\
A_{5n} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{s}^{-i}[\theta_0](x_i) - f_0^{[1]}(x_i, y_i)}{f_0^{[0]}(x_i)} (f_0^{[0]}(x_i) - \hat{\tau}^{-i}(x_i, y_i)) \\
A_{6n} &= \frac{1}{n} \sum_{i=1}^{n} \nabla_0 \log(\hat{\tau}^{-i}(\theta_0(x_i, y_i))) \bigg|_{\theta_0} - \frac{f_0^{[0]}(x_i) - \hat{\tau}^{-i}(x_i, y_i)}{f_0^{[0]}(x_i)} \quad \frac{(f_0^{[0]}(x_i) - \hat{\tau}^{-i}(x_i, y_i))^2}{f_0^{[0]}(x_i)} \\
A_{7n} &= \frac{1}{n} \sum_{i=1}^{n} \frac{f_0^{[1]}(x_i) - \hat{s}^{-i}[\theta_0](x_i)}{f_0^{[0]}(x_i)} (\hat{s}^{-i}(x_i, y_i) - f_0^{[0]}(x_i)) \\
A_{8n} &= \frac{1}{n} \sum_{i=1}^{n} \nabla_0 \log(\hat{s}^{-i}(\theta_0(x_i))) \bigg|_{\theta_0} \frac{(f_0^{[0]}(x_i) - \hat{s}^{-i}(x_i))^2}{f_0^{[0]}(x_i)} 
\end{align*}
\]
To deal with each of these terms separately, let us define the following functions:

\[
\tilde{r}(\theta', x, y) = \frac{1}{h_1 h_2} \int K\left(\frac{\theta'(x - u)}{h_1}\right) K\left(\frac{y - v}{h_2}\right) f(u, v) du dv,
\]

\[
\tilde{s}(\theta', x) = \frac{1}{h_1} \int K\left(\frac{\theta'(x - u)}{h_1}\right) f(u, v) du dv
\]

For instance, to deal with \( A_{1n} \) we telescope using \( r^{[1]}(\theta'_0 X_i, Y_i) \) to obtain \( A_{1n} = B_{1n} + C_{1n} \), where

\[
B_{1n} = \frac{1}{n} \sum_{i=1}^{n} \frac{\tilde{r}_i^{[1]}(\theta'_0 X_i, Y_i)}{f_{\theta_0}(\theta'_0 X_i, Y_i)}
\]

\[
C_{1n} = \frac{1}{n} \sum_{i=1}^{n} \frac{\tilde{r}_i^{[1]}(\theta'_0 X_i, Y_i) - \tilde{r}_i^{[1]}(\theta'_0 X_i, Y_i)}{f_{\theta_0}(\theta'_0 X_i, Y_i)}
\]

and \( A_{2n} = B_{2n} + C_{2n} \), where

\[
B_{2n} = \frac{1}{n} \sum_{i=1}^{n} \frac{f_{\theta_0}^{[1]}(\theta'_0 X_i, Y_i) \left( \tilde{r}_i^{[1]}(\theta'_0 X_i, Y_i) - \tilde{r}(\theta'_0 X_i, Y_i) \right)}{f_{\theta_0}(\theta'_0 X_i, Y_i)}
\]

\[
C_{2n} = \frac{1}{n} \sum_{i=1}^{n} \frac{f_{\theta_0}^{[1]}(\theta'_0 X_i, Y_i) f_{\theta_0}(\theta'_0 X_i, Y_i) \left( \tilde{r}(\theta'_0 X_i, Y_i) - \tilde{r}(\theta'_0 X_i, Y_i) \right)}{f_{\theta_0}(\theta'_0 X_i, Y_i)}
\]

In a parallel way \( A_{in} = B_{in} + C_{in} \) for \( i = 3, 4 \). Just defining \( B_n = B_{1n} + B_{2n} + B_{3n} + B_{4n} \), \( C_n = C_{1n} + C_{2n} + C_{3n} + C_{4n} \) and \( D_n = A_{5n} + A_{6n} + A_{7n} + A_{8n} \) and using equation (21) we have

\[
\alpha_0(\theta_0) = B_n + C_n + D_n + o_p(n^{-1/2}).
\]

Now using Lemmas 5, 8 and 9 implies that

\[
\sqrt{n} C_n = o_p(1), \quad \sqrt{n} B_n = o_p(1), \quad \sqrt{n} D_n = o_p(1).
\]

Finally, Lemma 10 can be used to prove \( l^{[2]}_m(\theta) \to l^{[2]}(\theta) \). This completes the proof. □

Next we consider the gradients of \( f_{\theta} \). We have the following results:

**Lemma 3.** Under A6, we have
\[ \frac{\partial}{\partial \theta_k} f_\delta(\theta' x, y)_{\theta = \theta_0} = \frac{\partial}{\partial t} \left( f_{\theta_0}(t, y) \left[ x_k - E(X_k | \theta_0' X = t, Y = y) \right] \right)_{t = \theta_0' x}. \]

\[ \frac{\partial}{\partial \theta_k} f_\delta(\theta' x, y)_{\theta = \theta_0} = \frac{\partial}{\partial t} \left( f_{\theta_0}(t, y) \left[ x_k - E(X_k | \theta_0' X = t) \right] \right)_{t = \theta_0' x}. \]

**Proof.** We will prove only part a). The proof of b) is very similar. Let us consider \( \theta_0 = (\theta_{01}, ..., \theta_{0d})' \) and \( \theta_0(k) = \theta_0 + (0, ..., \delta, ..., 0)' \), where the \( \delta \) is in position \( k \). Then, for every \( x = (x_1, ..., x_d)' \) and \( y \in \mathbb{R} \), we have

\[ (26) \quad \frac{\partial}{\partial \theta_k} f_\delta(\theta' x, y)_{\theta = \theta_0} = \beta_1 + \beta_2 + \beta_3, \]

where

\[ (27) \quad \beta_1 = \lim_{\delta \to 0} \frac{f_{\theta_0(k)}(\theta_0' x, y) - f_{\theta_0}(\theta_0' x, y)}{\delta} \]

\[ (28) \quad \beta_2 = \lim_{\delta \to 0} \frac{f_{\theta_0}(\theta_0' x, y) - f_{\theta_0}(\theta_0' x, y)}{\delta} \]

\[ (29) \quad \beta_3 = \lim_{\delta \to 0} \left[ \frac{f_{\theta_0(k)'}(\theta_0' x, y) - f_{\theta_0}(\theta_0' x, y)}{\delta} - \frac{f_{\theta_0(k)}(\theta_0' x, y) - f_{\theta_0}(\theta_0' x, y)}{\delta} \right]. \]

Using two Taylor expansions, the term in (29) becomes

\[ \beta_3 = \lim_{\delta \to 0} \left[ \frac{\partial}{\partial \theta_k} f_\delta(\theta_0' x, y)_{\theta = \theta_0} - \frac{\partial^2}{\partial \theta_k^2} f_\delta(\theta_0' x, y)_{\theta = \tilde{\theta}} \right] = 0 \]

where \( \tilde{\theta} = \theta_0 + (0, ..., \delta, ..., 0)' \) and \( \tilde{\delta} \) and \( \tilde{\delta} \) are intermediate points between 0 and \( \delta \) and the last step comes from the continuity of the first and second partial derivatives (see Condition A6).

A direct inspection of (28) leads to

\[ \beta_2 = \frac{\partial}{\partial \theta_k} f_\delta(\theta' x, y)_{\theta = \theta_0} = x_k \frac{\partial}{\partial t} f_\delta(t, y)_{t = \theta_0' x}. \]
On the other hand

\begin{equation}
\beta_1 = \frac{\partial}{\partial \theta_k} f_\theta(\theta'_0 x, y)_{\theta=\theta_0} \\
= \frac{\partial}{\partial u} \left( \frac{\partial}{\partial v} \left( \frac{\partial}{\partial \theta_k} P(\theta' X \leq u, Y \leq v)_{\theta=\theta_0} \right)_{v=y} \right)_{u=\theta'_0 x}.
\end{equation}

Let \( f_{X_k, \theta'_0 X, Y}(x_k, u, y) \) denote the density of \((X_k, \theta'_0 X, Y)\). Now using standard algebra for the inner partial derivative in (30), we obtain

\begin{align*}
\frac{\partial}{\partial \theta_k} P(\theta' X \leq u, Y \leq v)_{\theta=\theta_0} &= \\
= \lim_{\delta \to 0} \frac{P(\theta'_0 X + \delta X_k \leq u, Y \leq v) - P(\theta'_0 X \leq u, Y \leq v)}{\delta} \\
(31) &= \int_{-\infty}^{u} \int_{-\infty}^{\infty} \lim_{\delta \to 0} \frac{1}{\delta} \left[ \int_{-\infty}^{u-\delta z_1} f_{X_k, \theta'_0 X, Y}(z_1, z_2, z_3) dz_2 \\
&\quad - \int_{-\infty}^{u} f_{X_k, \theta'_0 X, Y}(z_1, z_2, z_3) dz_2 \right] dz_1 dz_3 \\
&\quad - \int_{-\infty}^{u} \int_{-\infty}^{\infty} z_1 f_{X_k, \theta'_0 X, Y}(z_1, u, z_3) dz_1 dz_3.
\end{align*}

Now using (31) in (30) gives

\begin{equation}
\beta_1 = -\frac{\partial}{\partial u} \left( f_{\theta_0}(u, y)E(X_k|\theta'_0 X = u, Y = y)_{\theta=\theta_0}. \right.
\end{equation}

Using (29), (28) and (32) in (26) the proof of a) is concluded. \( \square \)

**Lemma 4.** Under A2, we have

\begin{align*}
a) & \quad \nabla_\theta f_\theta(y|\theta' x)_{\theta=\theta_0} = \left[ x - E(X|\theta'_0 X = \theta'_0 x) \right] \frac{\partial}{\partial \theta_0} f_{\theta_0}(y|\theta_0)_{\theta=\theta_0}. \\
\quad b) & \quad E(\nabla_\theta \log f_\theta(Y_i|\theta'_0 X_i))_{\theta=\theta_0} = 0
\end{align*}

**Proof.** Part a) is an immediate consequence of Lemma 3. For part b), A2 implies

\begin{align*}
E \left[ |X_i - E(X|\theta'_0 X = \theta'_0 X_i)|Y_i, \theta'_0 X_i \right] \\
= E(X|\theta'_0 X = \theta'_0 X_i) - E(X|\theta'_0 X = \theta'_0 X_i) = 0.
\end{align*}
LEMMA 5. Under A2, A4 and A5, \( \sqrt{n}C_n = O_P(\sqrt{n}h_4^4 + \sqrt{n}h_2^2) = o_P(1) \).

Proof. First, using several changes of variables, we have

\[
\hat{r}(\theta_0'X_i, Y_i)1_{\{Y_i \leq \alpha_n\}} = \frac{1}{h_1h_2} \int \int K \left( \frac{\theta_0'X_i - u}{h_1} \right) K \left( \frac{Y_i - v}{h_2} \right) f_{\theta_0}(u, v)du dv = \int \int K(z_1) K(z_2) f_{\theta_0}(\theta_0'X_i - h_1z_1, Y_i - h_2z_2)dz_1 dz_2
\]

and

\[
\hat{r}^{[1]}(\theta_0'X_i, Y_i)1_{\{Y_i \leq \alpha_n\}} = \frac{1}{h_1h_2} \int f(X_i - E(X|\theta_0'X = u))K' \left( \frac{\theta_0'X_i - u}{h_1} \right) f_{\theta_0}(u, v)du dv = \frac{1}{h_1} \int f(X_i - E(X|\theta_0'X = \theta_0'X_i - h_1z_1))K'(z_1)K(z_2) f_{\theta_0}(\theta_0'X_i - h_1z_1, Y_i - h_2z_2)dz_1 dz_2.
\]

Using A5, the function \( f_\theta(u, v) \) is three times differentiable. Now applying a Taylor expansion and using \( \int z^aK(z)dz = 0 \) for \( a \) odd, \( \int K(z)dz = 1 \) and \( d_K = \int z^2K(z)dz < \infty \), together with \( \int z^bK'(z)dz = 0 \) for \( b \) even, \( \int zK'(z)dz = -1 \) and \( \int z^3K'(z)dz = -3 \int z^2K(z)dz \) we obtain

\[
\hat{r}(\theta_0'X_i, Y_i) = f_{\theta_0}(\theta_0'X_i, Y_i) + \frac{d_K h_1^2}{2} \frac{\partial^2}{\partial^2 u} f_{\theta_0}(u, Y_i)_{u=\theta_0'X_i}
\]

\[(33) + \frac{d_K h_1^2}{2} \frac{\partial^2}{\partial^2 v} (f_{\theta_0}(\theta_0'X_i, v)1_{\{Y_i \leq \alpha_n\}})_{v=Y_i} + O_P(h_1^4 + h_1^2h_2^2 + h_2^4).\]

Similarly, Lemma 3 implies that

\[
\hat{r}^{[1]}(\theta_0'X_i, Y_i) = f_{\theta_0}(\theta_0'X_i, Y_i) + \frac{d_K h_1^2}{2} \frac{\partial^2}{\partial^2 u} [(X_i - E(X|\theta_0'X = u))f_{\theta_0}(u, Y_i)]_{u=\theta_0'X_i}
\]

\[+ \frac{d_K h_2^2}{2} \frac{\partial}{\partial u} \frac{\partial^2}{\partial^2 v} [(X_i - E(X|\theta_0'X = u))f_{\theta_0}(u, v)]_{u=\theta_0'X_i, v=Y_i}
\]

\[(34) + O_P(h_1^4 + h_1^2h_2^2 + h_2^4).\]

Now starting from (24), using (33) and (34) and repeating similar steps for \( \hat{s}(\theta_0'X_i) \), we obtain

\[
C_n = \hat{C}_{1n} + \hat{C}_{2n} + O_P(h_1^4 + h_2^4),
\]

\[(35)\]
where

\[ \hat{C}_{1n} = \frac{1}{n} \sum_{i=1}^{n} \frac{dK h_i^2}{2} \hat{C}_{1i}, \]

\[ \hat{C}_{2n} = \frac{1}{n} \sum_{i=1}^{n} \frac{dK h_i^2}{2} \hat{C}_{2i}, \]

\[ \hat{C}_{1i} = \frac{1}{f_{00}(\theta_{0i}^0 Y_i, \theta_{0i}^0 X_i)} \frac{d^3}{du^3} \left\{ \left[ X_i - E(X|\theta_{0i}^0 X = u) \right] f_{00}(u, Y_i) \right\}_{u = \theta_{0i}^0 X_i} \]

\[ - \frac{1}{f_{00}(\theta_{0i}^0 X_i)} \frac{d^3}{du^3} \left\{ \left[ X_i - E(X|\theta_{0i}^0 X = u) \right] f_{00}(u, X_i) \right\}_{u = \theta_{0i}^0 X_i} \]

and

\[ \hat{C}_{2i} = - \frac{\nabla \theta_{0i} f_{00}\theta_{0i}^0 Y_i, \theta_{0i}^0 X_i} {f_{00}(\theta_{0i}^0 X_i)} \frac{d^2}{du^2} f_{00}(u, Y_i)_{u = \theta_{0i}^0 X_i} + \nabla \theta_{0i} f_{00}\theta_{0i}^0 Y_i, \theta_{0i}^0 X_i \frac{d}{du} f_{00}(u, X_i)_{u = \theta_{0i}^0 X_i}. \]

We show that \( E(\hat{C}_{kn}) = 0 \) and \( Var(\hat{C}_{kn}) = o(n^{-1}) \) for \( k = 1, 2 \). Since \( f_{00}(z|\theta_{0i}^0 x) = f(z|x) \), we have that \( f(x, z) = f_{00}(\theta_{0i}^0 x, z) f_X(x)/f_{00}(\theta_{0i}^0 x) \).

This equation and condition A2 can be used to obtain:

\[ E(\hat{C}_{1i}) = \int \frac{1}{f_{00}(\theta_{0i}^0 x, z)} \frac{d^3}{du^3} \left\{ \left[ x - E(X|\theta_{0i}^0 X = u) \right] f_{00}(u, y) \right\}_{u = \theta_{0i}^0 X_i} f(x) f(y) dx \ dy \]

\[ - \int \frac{1}{f_{00}(\theta_{0i}^0 x)} \frac{d^3}{du^3} \left\{ \left[ x - E(X|\theta_{0i}^0 X = u) \right] f_{00}(u, x) \right\}_{u = \theta_{0i}^0 X_i} f(x) dx = 0. \]

Moreover, the \( \hat{C}_{1i} \) are i.i.d. and, since \( h_i^2 \rightarrow 0 \), \( Var(\sqrt{n} \hat{C}_{1n}) \rightarrow 0 \). Hence \( \sqrt{n} \hat{C}_{1n} \rightarrow 0 \) in probability. Similar arguments can be used to conclude \( \sqrt{n} \hat{C}_{2n} \rightarrow 0 \). Now using (35), we have \( \sqrt{n} C_n = O_P(h_1^2 + \sqrt{n}(h_1^4 + h_2^2)) = o_P(1) \), which completes the proof.

**Lemma 6.** Under the conditions in Theorem 3 we have \( B_{1n} = o_P(n^{-1/2}) \).

**Proof.** Starting from (23) and using (22) and (7) we obtain

\[ B_{1n} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{f_{00}(\theta_{0i}^0 X_i, Y_i)} \left[ \frac{1}{h_1^2 h_2 (n - 1)} \sum_{j \neq i}^n (X_i - X_j) K' \left( \frac{\theta_{0i}^0 X_i - \theta_{0j}^0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \right] \]

\[ = \frac{1}{h_1^2 h_2(n - 1)} \int \int (X_i - u) K' \left( \frac{\theta_{0i}^0 X_i - \theta_{0j}^0 u}{h_1} \right) K \left( \frac{Y_i - u}{h_2} \right) f(u, v) \ dudv \]

\[ = \frac{1}{n(n - 1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} H_{ij}, \]
where \(K'(t)\) denotes the derivative of \(K\) with respect to \(t\) and

\[
H_{ij} = \frac{1}{h_1^2 h_2} \int_0^1 \left[ (X_i - X_j)K' \left( \frac{\theta'_0 X_i - \theta'_0 X_j}{h_1} \right) K \left( \frac{Y_i - Y_j}{h_2} \right) \right]
- \int \left( X_i - u \right)K' \left( \frac{\theta'_0 X_i - \theta'_0 u}{h_1} \right) K \left( \frac{Y_i - u}{h_2} \right) f(u, v) \, du \, dv.
\]

Let us define

\[
\bar{H}_j = E(H_{ij} | X_j, Y_j, \delta_j).
\]

Then \(B_{1n} = B_{1n1} + B_{1n2}\), where

\[
\begin{align*}
B_{1n1} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} (H_{ij} - \bar{H}_j) \\
B_{1n2} &= \frac{1}{n} \sum_{j=1}^{n} \bar{H}_j.
\end{align*}
\]

For the term in (36) it is evident that

\[
E(H_{ij} - \bar{H}_j) = E(H_{ij} - \bar{H}_j | X_i, Y_i) = E(H_{ij} - \bar{H}_j | X_j, Y_j) = 0,
\]

so \(E(B_{1n1}) = 0\). On the other hand, long but straightforward calculations can be performed to compute the variance of \(B_{1n1}^{(m)}\), the \(m\)-th component of the random vector \(B_{1n1}\), which results:

\[
\text{Var}(B_{1n1}^{(m)}) = E(B_{1n1}^{(m)})^2 = \frac{1}{n^2(n-1)^2} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \sum_{k=1}^{n} \sum_{l=1, l \neq k}^{n} E((H_{ij}^{(m)} - \bar{H}_j^{(m)})(H_{kl}^{(m)} - \bar{H}_l^{(m)}))
= \frac{1}{n(n-1)} [E((H_{12}^{(m)} - \bar{H}_2^{(m)})^2) + E((H_{12}^{(m)} - \bar{H}_2^{(m)})(H_{21}^{(m)} - \bar{H}_1^{(m)}))].
\]

Using standard algebra, the last two expectations can be proved to be of order \(O(h_1^{-3}h_2^{-1})\). As a consequence \(\text{Var}(B_{1n1}^{(m)}) = O(n^{-2}h_1^{-3}h_2^{-1})\). Now Condition A4 implies \(B_{1n1}^{(m)} = o_P(n^{-1/2})\). The term in (37) can be handled in a similar way to prove \(E(B_{1n2}) = 0\) and \(\text{Var}(B_{1n2}^{(m)}) = O(n^{-1}h_1)\). Thus, A4 implies \(B_{1n2}^{(m)} = o_P(n^{-1/2})\). This concludes the proof.

Now we state a similar lemma.
LEMMA 7. Under the conditions in Theorem 3 we have \( B_{2n} = o_p(n^{-1/2}) \), \( B_{3n} = o_p(n^{-1/2}) \) and \( B_{4n} = o_p(n^{-1/2}) \).

**Proof.** Given the similarities between the term \( B_{1n} \) in (23) and the term \( B_{2n} \) in (25), and also the terms \( B_{3n} \) and \( B_{4n} \), the proof is parallel to that of Lemma 6, obtaining \( B_{2n} = B_{2n1} + B_{2n2} \), \( B_{3n} = B_{3n1} + B_{3n2} \), \( B_{4n} = B_{4n1} + B_{4n2} \), where \( E[B_{jnk}] = 0 \), for \( j = 2, 3, 4 \) and \( k = 1, 2 \), and \( \text{Var}(B_{2n1}) = O(n^{-2}h_1^{-1}h_2^{-1}) \), \( \text{Var}(B_{2n2}) = O(n^{-1}h_2^2) \), \( \text{Var}(B_{3n1}) = O(n^{-2}h_1^{-3}) \), \( \text{Var}(B_{3n2}) = O(n^{-1}h_1) \), \( \text{Var}(B_{4n1}) = O(n^{-2}h_1^{-1}) \), \( \text{Var}(B_{4n2}) = O(n^{-1}h_2^2) \). Now assumption A4 implies that the orders in the six variance terms are all \( o(n^{-1}) \), which concludes the proof. \( \square \)

**LEMMA 8.** Under the conditions in Theorem 3, \( B_n = o_p(n^{-1/2}) \).

**Proof.** The proof makes use of the definition of \( B_n = B_{1n} + B_{2n} + B_{3n} + B_{4n} \) and Lemmas 6 and 7. \( \square \)

**LEMMA 9.** Under the conditions in Theorem 3, \( D_n = o_p(n^{-1/2}) \).

**Proof.** Each of the terms \( A_{5n} \), \( A_{6n} \), \( A_{7n} \), \( A_{8n} \) can be bounded using the Cauchy-Schwarz inequality. Let us consider \( A_{5n} \):

\[
|A_{5n}| \leq A_{5n1}^{1/2} A_{5n2}^{1/2},
\]

where

\[
A_{5n1} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{f}^{-1}(\theta_0'X_i, Y_i) - f_{\theta_0}^{[1]}(\theta_0'X_i, Y_i))^2}{f_{\theta_0}^{[1]}(\theta_0'X_i, Y_i)}
\]

\[
A_{5n2} = \frac{1}{n} \sum_{i=1}^{n} \frac{(f_{\theta_0}^{[1]}(\theta_0'X_i, Y_i) - \hat{f}^{-1}(\theta_0'X_i, Y_i))^2}{f_{\theta_0}^{[1]}(\theta_0'X_i, Y_i)}
\]

These two terms can be expanded as

\[
A_{5n1} = A_{5n11} + A_{5n12} + A_{5n13},
\]

\[
A_{5n2} = A_{5n21} + A_{5n22} + A_{5n23},
\]

(38) \begin{align*}
A_{5n1} &= A_{5n11} + A_{5n12} + A_{5n13}, \\
A_{5n2} &= A_{5n21} + A_{5n22} + A_{5n23},
\end{align*}
where

\[
A_{5n11} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{r}^{-i}(\theta_0')X_i, Y_i) - \hat{r}^{[1]}(\theta_0'X_i, Y_i))^2}{f_{\theta_0}^2(\theta_0'X_i, Y_i)}
\]

\[
A_{5n12} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{r}^{[1]}(\theta_0'X_i, Y_i) - f_{\theta_0}^{[1]}(\theta_0'X_i, Y_i))^2}{f_{\theta_0}^2(\theta_0'X_i, Y_i)}
\]

\[
A_{5n13} = \frac{2}{n} \sum_{i=1}^{n} \frac{((\hat{r}^{-i}(\theta_0'X_i, Y_i) - \hat{r}^{[1]}(\theta_0'X_i, Y_i))(\hat{r}^{[1]}(\theta_0'X_i, Y_i) - f_{\theta_0}^{[1]}(\theta_0'X_i, Y_i))}{f_{\theta_0}^2(\theta_0'X_i, Y_i)}
\]

\[
A_{5n21} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{r}_0(\theta_0'X_i, Y_i) - \hat{r}(\theta_0'X_i, Y_i))^2}{f_{\theta_0}^2(\theta_0'X_i, Y_i)}
\]

\[
A_{5n22} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{r}(\theta_0'X_i, Y_i) - \hat{r}^{-i}(\theta_0'X_i, Y_i))^2}{f_{\theta_0}^2(\theta_0'X_i, Y_i)}
\]

\[
A_{5n23} = \frac{2}{n} \sum_{i=1}^{n} \frac{(\hat{r}_0(\theta_0'X_i, Y_i) - \hat{r}(\theta_0'X_i, Y_i))(\hat{r}(\theta_0'X_i, Y_i) - \hat{r}^{-i}(\theta_0'X_i, Y_i))}{f_{\theta_0}^2(\theta_0'X_i, Y_i)}
\]

The terms \(A_{5n11}, A_{5n12}, A_{5n21}\) and \(A_{5n22}\) can be treated similarly to \(B_{1n}, C_{1n}, B_{2n}, C_{2n}\), respectively. On the other hand, \(A_{5n13}\) (respectively \(A_{5n23}\)) can be bounded, using the Cauchy-Schwarz inequality, in terms of \(A_{5n11}\) and \(A_{5n12}\) (respectively \(A_{5n21}\) and \(A_{5n22}\)). All in all, the results in \(A_{5nk}\) for \(j = 1, 2\) and \(k = 1, 2\). In view of (39) (40) we have \(A_{5n1} = o_P(n^{-1/2})\) and \(A_{5n2} = o_P(n^{-1/2})\), which, using (38), gives \(A_{5n} = o_P(n^{-1/2})\).

Following parallel arguments, straightforward but tedious algebra can be used to prove that \(A_{6n} = o_P(n^{-1/2}), A_{7n} = o_P(n^{-1/2})\) and \(A_{8n} = o_P(n^{-1/2})\).

Since \(D_n = A_{5n} + A_{6n} + A_{7n} + A_{8n}\), this concludes the proof.

**Lemma 10.** Under the conditions in Theorem 3, \(\hat{i}^{[2]}_n(\theta) \rightarrow i^{[2]}(\theta)\).

**Proof.** We have

\[
\hat{i}^{[2]}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{r}^{-i}(\theta'X_i, Y_i) - \hat{r}^{[1]}(\theta'X_i, Y_i)(\hat{r}^{-i}(\theta'X_i, Y_i))'}{\hat{r}^{-i}(\theta'X_i, Y_i)^2} - \frac{\hat{s}^{-i}(\theta'X_i)}{\hat{s}^{-i}(\theta'X_i)^2}
\]

\[+ \frac{\hat{r}^{-i}(\theta'X_i)(\hat{s}^{-i}(\theta'X_i))'}{\hat{s}^{-i}(\theta'X_i)^2}
\]
Since
\[ \hat{\tau}^{-1}\{y, \theta'x\} = \frac{1}{nh_1h_2} \sum_{j=1}^{n} (x - X_j)(x - X_j)'K'' \left( \frac{\theta'x - \theta'X_j}{h_1} \right) K \left( \frac{y - Y_j}{h_2} \right), \]

it is easy to show, that
\[ \hat{\tau}^{-1}\{\theta'x, y\} \to \frac{\partial^2}{\partial u^2} \{f_\theta(u, y)E(\langle x - X\rangle\langle x - X\rangle'|\theta'X = u)\}_{u=\theta'x} \text{ in probability.} \]

Moreover, using Lemma 3, it can be proved that
\[ \frac{\partial^2}{\partial u^2} \{f_\theta(u, u)E(\langle x - X\rangle\langle x - X\rangle'|\theta'X = u)\}_{u=\theta'x} = f^{[2]}_\theta(\theta'x, y). \]

Similarly, it can be shown that \( \hat{s}^{[2]}(\theta'x) \to f^{[2]}_\theta(\theta'x) \) in probability. This completes the proof. \( \square \)

**Proof of Theorem 2.** In view of Lemmas 1, 2 and 10, it remains to be shown that \( \sum_{k=1}^{8} A_{kn} \overset{P}{\to} 0. \) But this can be proved in line with Lemmas 5, 8 and 9 but even more simply, since only convergence in probability to zero (and no rate) is required. \( \square \)