SUPPLEMENT: PROOFS OF THEOREMS 2.2 AND 2.3

For a matrix $M$, let $||M|| = \max_{i,j} |m_{ij}|$. It is easy to see that

**Lemma A.1** A sequence of square random matrices $W_n$ converges to $W$ in probability if and only if $a^T W_n a \to a^T W a$ in probability for all $a \in \mathbb{R}^p$.

**Proof of Theorem 2.2**: We prove the results for $U$-statistic $S$ in (1) of any degree $K \geq 1$. For any $a \in \mathbb{R}^p$, let $S^\dagger(a) = a^T S(a)$. The Sen (1960) variance estimator of $S^\dagger(a)$ is $V^\dagger(a) = a^T V(a) a$ where $V$ is defined in (5). From Sen (1960), it can be seen that $nK^{-2} V^\dagger(a) \to a^T \zeta_1 a$ in probability. From Lemma A.1, we obtain (7) and (8) follows from Slutsky’s theorem and Theorem 2.1.

In preparation for Theorem 2.3, let

$$A_n(\theta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k,k \neq j} h(Z_i, Z_j; \theta) h^T(Z_i, Z_k; \theta),$$

$$B_n(\theta) = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k,k \neq j} h(Z_i, Z_j; \theta) h^T(Z_i, Z_k; \theta_0),$$

$$\tilde{\zeta}_2(\theta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n h(\otimes_2(Z_i, Z_j; \theta),$$

$$C_n(\theta) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n h(Z_i, Z_j; \theta) h^T(Z_i, Z_j; \theta_0).$$

The following assumptions are used to establish the results of Theorem 2.3.

**Assumptions:**

1. The true value of $\theta$ is $\theta_0$, an interior point of the parameter space $\Omega$. Further, $E_{\theta_0}\{S(\theta)\} = 0$ has the unique solution $\theta_0$ and $\hat{\theta} \to \theta_0$ in probability.

2. $||\zeta_2|| < \infty$ and the determinant $|\zeta_1| > 0$.

3. $A_n(\hat{\theta}), A_n(\theta_0), B_n(\hat{\theta})$, and $B_n(\theta_0)$ converge in probability to $\zeta_1$. $\tilde{\zeta}_2(\hat{\theta}), \tilde{\zeta}_2(\theta_0), C_n(\hat{\theta})$, and $C_n(\theta_0)$ converge in probability to $\zeta_2$.

4. There exists a function $f(Z_1, Z_2)$ with $E\{f(Z_1, Z_2)\} < \infty$ and a neighborhood $B(\theta_0)$ of $\theta_0$ such that $\sup_{\theta \in B(\theta_0)} ||\tilde{\zeta}_2(\theta)|| \leq f(Z_1, Z_2)$. 

1
5. For almost every sequence of random vectors \(Z_1, Z_2, \ldots, Z_n, \ldots\) in \(\mathbb{R}^p\) and any \(\epsilon > 0\),
\[
\sum_{i=1}^{n} ||T_{n,i}||^2 I\{||T_{n,i}|| > \epsilon\} \to 0,
\]
where \(T_{n,i} = \hat{\zeta}_1(\hat{\theta})^{-1/2} \sum_{j=1}^{n} h(Z_i, Z_j; \hat{\theta})/n\) and \(\hat{\zeta}_1(\hat{\theta}) = n^{-2}(n - 1)^3V(\hat{\theta})/4\).

The Assumption 3 is needed to be able to substitute \(\hat{\theta}\) into the estimating function before bootstrapping the terms. It is of some interest to note that these results would follow from conditions of uniform convergence. For example, if \(A_n(\theta)\) converges to its probability limit \(\zeta_1(\theta, \theta_0) = E_{\theta_0}\{h(Z_1, Z_2; \theta)h^T(Z_1, Z_3; \theta)\}\) uniformly in a neighborhood of \(\theta_0\) and \(\zeta_1(\theta, \theta_0)\) is continuous in \(\theta\) for \(\theta \in \mathcal{B}(\theta_0)\), then \(A_n(\theta_0) \to \zeta_1\) and \(A_n(\hat{\theta}) \to \zeta_1\) in probability.

**Proof of Theorem 2.3:** Some approaches in the following proof derive from Theorems 1 and 4 of Athreya et al. (1984) and Theorem 7.1 of Hoeffding (1948).

Let \(F_n\) be the empirical distribution function of \(Z_1, \ldots, Z_n\), and let \(Z^*_1, \ldots, Z^*_n\) be a random sample from \(F_n\). The expressions (5) and (6) for \(K = 2\) can be written as
\[
V^*(\hat{\theta}) = \frac{4}{n(n-1)} \mathbb{E} \left\{ \frac{1}{n-1} \sum_{j, j \neq i} h(Z^*_i, Z^*_j; \hat{\theta}) - S^*(\hat{\theta}) \right\} \otimes^2
\]
and \(S^*_i(\hat{\theta}) = V^*(\hat{\theta})^{-1/2} S^*(\hat{\theta})\) where \(S^*(\hat{\theta}) = 2n^{-1}(n-1)^{-1} \sum_{1 \leq i < j \leq n} h(Z^*_i, Z^*_j; \hat{\theta})\).

Let \(E^*(\cdot) = E(\cdot | F_n)\) and define \(h^*(z_1; \theta) = E^*\{h(z_1, Z^*_2; \theta)\} = n^{-1} \sum_{i=1}^{n} h(z^*_1, Z^*_i; \theta)\). Given \(F_n, h^*_1(Z^*_1; \hat{\theta}), \ldots, h^*_1(Z^*_n; \hat{\theta})\) are independent and identically distributed with mean 0 and conditional covariance matrix \(E^*\{h^*_1(\otimes^2; \hat{\theta})\} = \tilde{\zeta}_1(\hat{\theta})\) where we define
\[
\tilde{\zeta}_1(\theta) = \mathbb{E} \left\{ \frac{1}{n^3} \sum_{i_1=1}^{n} \left\{ \sum_{i_2=1}^{n} h(Z_{i_1}, Z_{i_2}; \theta) \right\} \right\} \otimes^2
\]

To show (11), consider first the case \(p = 1\) for which
\[
nV^*(\hat{\theta}) = \frac{4}{n-1} \sum_{i=1}^{n} b_i^2(\hat{\theta}) - \frac{4n}{n-1} S^2(\hat{\theta}) \tag{A.1}
\]
where \(b_i(\hat{\theta}) = (n - 1)^{-1} \sum_{j, j \neq i} h(Z^*_i, Z^*_j; \hat{\theta})\).

Given \(F_n, S^*(\hat{\theta})\) is a \(U\)-statistic and from (3.17) of Athreya et al. (1984) we obtain,
\[
E\{S^*(\hat{\theta})^2\} = E[E^*\{S^*(\hat{\theta})^2\}] = \left(\begin{array}{c} n \\ 2 \end{array}\right)^{-1} [2(n - 2)E\{\tilde{\zeta}_1(\hat{\theta})\} + E\{\tilde{\zeta}_2(\hat{\theta})\}] \leq \frac{4}{n} E\{\tilde{\zeta}_2(\hat{\theta})\}. \tag{A.2}
\]

By Assumptions 3, 4 and the dominated convergence theorem (Loève, 1977), \(E\{S^*(\hat{\theta})^2\} \to 0\) so that \(S^*(\hat{\theta}) \to 0\) in probability as \(n \to \infty\). In view of (A.1), it remains to show that
\[
\frac{1}{n} \sum_{i=1}^{n} \{b_i(\hat{\theta})\}^2 \to \zeta_1 \text{ in probability.} \tag{A.3}
\]
Thus, \( \frac{1}{n} \sum_{i=1}^{n} \{ b_i(\hat{\theta}) - h_i^*(Z_i^*; \hat{\theta}) \}^2 \to 0 \) in probability. According to Lemma 1 of Sen (1960) and (A.4), the result (A.3) follows if \( n^{-1} \sum_{i=1}^{n} h_1^2(Z_i^*; \hat{\theta}) \to \zeta_1 \) in probability. Following arguments for (3.24) and (3.25) in Athreya et al. (1984), it follows that \( n^{-1} \sum_{i=1}^{n} h_1^2(Z_i^*; \theta_0) \to \zeta_1 \) in probability, and it suffices to show that

\[
\frac{1}{n} \sum_{i=1}^{n} \{ h_i^*(Z_i^*; \hat{\theta}) - h_i^*(Z_i^*; \theta_0) \}^2 \to 0 \quad \text{in probability.} \tag{A.5}
\]

Given \( F_n \), the left side of (A.5) is an average of conditionally independent and identically distributed terms and has characteristic function \( \phi_n(t) = E\{\phi_n^*(t/n)\}^n \) where \( \phi_n^*(t/n) = E^*[\exp\{it(h_1^*(Z_i^*; \hat{\theta}) - h_1^*(Z_i^*; \theta_0))/n\}] \). To show (A.5), we show that \( \lim_{n \to \infty} \phi_n(t) = 1 \). Let \( r_n(t/n) = \int_{-\infty}^{\infty} \exp\{it(h_1^*(u; \hat{\theta}) - h_1^*(u; \theta_0))/n\} - 1 \)d\( F_n(u) \) so that \( \phi_n^*(t/n) = 1 + r_n(t/n) \). It suffices to show \( n|r_n(t/n)| \to 0 \) in probability. This follows in light of Assumption 3 and the fact that

\[
n|r_n(t/n)| \leq |t| \int_{-\infty}^{\infty} \{ h_i^*(u; \hat{\theta}) - h_i^*(u; \theta_0) \}^2 dF_n(u)
\]

\[
= \frac{|t|}{n} \sum_{i=1}^{n} \left\{ \frac{1}{n} \sum_{j=1}^{n} h(Z_i, Z_j; \hat{\theta}) - \frac{1}{n} \sum_{j=1}^{n} h(Z_i, Z_j; \theta_0) \right\}^2 \tag{A.6}
\]

\[
= \frac{|t|}{n} \{ \tilde{\zeta}_2(\hat{\theta}) + \tilde{\zeta}_2(\theta_0) - 2C_n(\hat{\theta}) \} + |t|\{A_n(\hat{\theta}) + A_n(\theta_0) - 2B_n(\hat{\theta})\}.
\]

This establishes (11) in the single parameter case \( (p = 1) \). To show (11) in the multiple parameter case, we replicate the above steps for the scalar-valued \( U \)-statistic \( a^T S(\theta) \) and use Lemma A.1.

To show (12), we first show that

\[
V(\hat{\theta})^{-1/2} S^*(\hat{\theta}) \to N(\theta, 1) \quad \text{in distribution.} \tag{A.7}
\]

Let \( X_{n,i} = \tilde{\zeta}_1(\hat{\theta})^{-1/2} h_1^*(Z_i^*; \hat{\theta}) \). For \( n = 1, 2, \ldots \) and \( i = 1, \ldots, n \), \( \{X_{n,i}\} \) is a triangular array whose elements in the \( n \)th row are independent and identically distributed conditional on \( F_n \). The corresponding Lindeberg condition is that

\[
\sum_{i=1}^{n} E^*||X_{n,i}||^2 I\{||X_{n,i}|| > \epsilon\} \to 0
\]

3
for all $\epsilon > 0$. This has the same form as Assumption 5 so that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{n,i} \rightarrow N(\theta, 1)$ in distribution.

Let $Y_n = (n - 1)^{3/2} n^{-1} S(Y)\theta$ and $Z_n = 2n^{-1/2} \sum_{i=1}^{n} h_i(Z_i; \hat{\theta})$. Now,

$$V(\hat{\theta})^{-\frac{1}{2}} S(Y)\theta - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{n,i} = \{4\zeta_i(\hat{\theta})\}^{-\frac{1}{2}} (Y_n - Z_n).$$

To show (A.7), we show that $\zeta_i(\hat{\theta}) = (n - 1)^{3/2} n^{-2} V(\hat{\theta})/4 \rightarrow \zeta_i$ and $Y_n - Z_n \rightarrow 0$ in probability.

By Theorem 2.2, $na^T V(\theta_0) a/4 \rightarrow a^T \zeta a$ in probability. Since $S(\theta_0) \rightarrow 0$ in probability and

$$na^T V(\theta_0) a = \frac{4}{n - 1} \sum_{i=1}^{n} \left\{ \frac{1}{n - 1} \sum_{j: j \neq i} a^T h(Z_i, Z_j; \theta) \right\}^2 - \frac{4n}{n - 1} \{a^T S(\theta)\}^2,$$

Lemma 1 of Sen (1960) and Lemma A.1 imply that $\zeta_i(\hat{\theta}) \rightarrow \zeta_i$ in probability if

$$na^T V(\theta) a = \frac{4}{n - 1} \sum_{i=1}^{n} \left\{ \frac{1}{n} \sum_{j=1}^{n} a^T h(Z_i, Z_j; \theta) - \frac{1}{n} \sum_{j=1}^{n} a^T h(Z_i, Z_j; \theta_0) \right\}^2 \rightarrow 0$$

in probability. This follows from a similar argument used to verify (A.6).

Finally, $Y_n - Z_n \rightarrow \theta$ in probability if $E(Y_n^{(l)} - Z_n^{(l)})^2 \rightarrow 0$ for each $l = 1, \ldots, p$, where $a^{(l)}$ stands for the $l$th component of a vector $a$. Given $F_n$, $S(Y)\theta$ is a $U$-statistic with mean $\theta$ so that

$$E^*(Y_n^{(l)})^2 = \frac{4(n - 1)^2(n - 2)}{n^3} \zeta_{1}^{(l)} + \frac{2(n - 1)^2}{n^3} \zeta_{2}^{(l)}.$$

Since $Z_n$ is a sum of conditional independent vectors, $E^*(Z_n^{(l)})^2 = 4 \zeta_{1}^{(l)}$. Moreover,

$$E^*(Y_n^{(l)} Z_n^{(l)}) = \frac{4}{n(n - 1)} \left( \frac{n - 1}{n} \right)^{\frac{3}{2}} \sum_{i=1}^{n} \sum_{1 \leq i_1 < i_2 \leq n} E^*\{h_i^*(Z_i^{*}; \hat{\theta}) h(Z_{i_1}^{*}, Z_{i_2}^{*}; \hat{\theta})\}$$

where the sum contains $n(n - 1)$ nonzero terms $E^*\{h_i^*(Z_i^{*}; \hat{\theta}) h(Z_{i_1}^{*}, Z_{i_2}^{*}; \hat{\theta})\} = \zeta_{1}^{(l)}$ when $i_1 = i$ or $i_2 = i$. It follows that

$$E(Y_n^{(l)} - Z_n^{(l)})^2 = E\{E^*(Y_n^{(l)} - Z_n^{(l)})^2\} \leq \frac{10}{n} E\{\zeta_{2}^{(l)}\},$$

and by an argument similar to that leading to (A.2), $E(Y_n^{(l)} - Z_n^{(l)})^2 \rightarrow 0$ which establishes (A.7).

The result (12) follows from Slutsky’s theorem.