ECO 519. Moment Maximal Inequalities for U-processes and Asymptotic Normality of Maximum Rank Correlation Estimator

This handout is a brief compendium of Professor Bob Sherman’s papers in Econometrica, Annals of Statistics and Econometric Theory cited in the list of readings.

U-Statistics and U-Processes
Let $P$ be a distribution on a set $S$, let $Z_1,\ldots,Z_n$ an iid sample from $P$. Let $f$ denote a real-valued function defined on $S^k = S \otimes S \cdot \cdot \cdot \otimes S$ with $k \geq 1$. We define the U-statistic of order $k$ by

$$U_{n,k} f = (n_k)^{-1} \sum_{i_k} f(Z_{i_1},\ldots,Z_{i_k})$$

where $(n)_k = n \times (n-1) \times \cdots \times (n-k+1)$, and $I_k$ is the set of all $(n)_k$ ordered $k$-tuples of distinct integers from the set $\{1,\ldots,n\}$. We will employ the following functional notation:

Take $k = 3$, then

$$f(P,s,t) = E[f(z_1, z_2, z_3) | z_2 = s, z_3 = t]; \quad f(P,s,P) = E[f(z_1, z_2, z_3) | z_2 = s]; \quad Qf = E[f(z_1, z_2, z_3)]$$

Note that $Q$ is the product measure $Q = P \otimes \cdots \otimes P$.

Suppose now that the function $f$ is such that under the product measure $Q = P \otimes \cdots \otimes P$, the conditional expectation of $f$ given any $k-1$ of its $k$ arguments is identically zero. Then we say that $f$ is $P$-degenerate, and that $U_{n,k} f$ is $P$-degenerate.

Hoeffding Decomposition
Let $f$, $P$ and $Q$ be as described above. If $Qf < \infty$, then there exist real-valued functions $f_1,\ldots,f_k$ such that for each $j$, $f_j$ is $P$-degenerate on $S_j$ and

$$U_{n,k} f = Qf + P_n f_1 + \sum_{j=2}^k U_{n,j} f_j$$

where, for each $z$ in $S$,

$$f_1(z) = f(z, P, \ldots, P) + f(P, z, P, \ldots, P) + \cdots + f(P, \ldots, P, z) - kQf$$
Moment Maximal Inequalities for U-Processes

In a completely analogous way to the one that yields moment maximal inequalities for Empirical Processes based on the properties of their packing and covering numbers, Bob Sherman characterized equivalent results for U-processes. We will only cite two corollaries of his main result here, which are used to prove the asymptotic normality of the MRC estimator:

**Lemma 1** Let $\mathcal{F}$ be a class of zero-mean functions $f$ on $S^k$, $k \geq 1$. If $\mathcal{F}$ is Euclidean for a constant envelope, then

$$\sup_{\mathcal{F}} |U_{n,k}| = O_p(1/\sqrt{n})$$

**Lemma 2** Let $\mathcal{F}$ be a class of $P$-degenerate functions on $S^k$, $k \geq 1$. If

(i) $\mathcal{F}$ contains the zero function.

(ii) $\mathcal{F}$ is Euclidean for the constant envelope $F$,

then

(a) $\sup_{\mathcal{F}} |n^{k/2}U_{n,k}f| = O_p(1)$.

(b) $\sup_{\mathcal{F}} |n^{k/2-\gamma}U_{n,k}f| \longrightarrow 0$ almost surely.

Heuristics of Asymptotic Normality of Maximum Rank Correlation (MRC) Estimator

The objective function is

$$G_n(\beta) = (n)^{-1} \sum_{i \neq j} 1\{Y_i > Y_j\} 1\{X_i'\beta > X_j'\beta\}$$

The maximizer is Han’s Maximum Rank Correlation (MRC) estimator. Proving consistency is relatively easy based on the assumptions:

(A1) The distribution of $X$ is continuous.

(A2) The function $F_0(\cdot)$ is strictly increasing in the support of $X'\beta_0$. 
The function \( G(\beta) = E[G_n(\beta)] \) is continuous everywhere in the parameter space.

Note that
\[
G(\beta) = E\left[1\{Y_i > Y_j\} 1\{X_i'\beta > X_j'\beta\}\right]
\]
\[
= E\left[\Pr[Y_i = 1, Y_j = 0|X_i, X_j] 1\{X_i'\beta > X_j'\beta\}\right]
\]
\[
= E\left[F_0(X_i'\beta_0)[1 - F_0(X_j'\beta_0)] 1\{X_i'\beta > X_j'\beta\}\right]
\]
\[
= \frac{1}{2} E\left[F_0(X_i'\beta_0)[1 - F_0(X_j'\beta_0)] 1\{X_i'\beta > X_j'\beta\} + F_0(X_j'\beta_0)[1 - F_0(X_i'\beta_0)] 1\{X_j'\beta > X_i'\beta\}\right]
\]

If we have \( \beta = \beta_0 \), then this becomes
\[
G(\beta_0) = \frac{1}{2} E\left[\max\{F_0(X_i'\beta_0)[1 - F_0(X_j'\beta_0)], F_0(X_j'\beta_0)[1 - F_0(X_i'\beta_0)]\}\right]
\]
So \( G(\beta) \) is clearly maximized at \( \beta = \beta_0 \). Assumptions (A1)-(A2) ensure that this is the unique maximizer.

To prove asymptotic normality, Sherman first re-expresses (symmetrizes) the objective function with the summands:
\[
\sum_{i<j} 1\{Y_i > Y_j\} 1\{X_i'\beta > X_j'\beta\} + 1\{Y_j > Y_i\} 1\{X_j'\beta > X_i'\beta\}
\]
define \( Z = (X, Y) \) and let
\[
\tau(z, \theta) = E[1\{y > Y\} 1\{x'\beta > x'\beta\}] + E[1\{Y > y\} 1\{X'\beta > x'\beta\}].
\]
Denote the normalized parameter vector by \( \theta \). Sherman chooses the normalization:
\[
\beta(\theta) = (\theta_1, \ldots, \theta_{d-1}, \sqrt{1 - \theta^2_1 - \cdots - \theta^2_{d-1}})
\]
(i.e, \( \|eta\| = 1 \)). Doing a switch of coordinates (easy), we can normalize the true parameter \( \theta_0 \) as \( \theta_0 = 0 \). Sherman shows that
\[
\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V^{-1}\Delta V^{-1}),
\]
where
\[
V = \frac{1}{2} E\left[\frac{\partial^2 \tau(Z, 0)}{\partial \theta \partial \theta'}\right], \quad \Delta = E\left[\frac{\partial \tau(Z, 0)}{\partial \theta} \frac{\partial \tau(Z, 0)'}{\partial \theta'}\right]
\]
The key is to show that the objective function can be expressed as:

\[ G_n(\theta) - G_n(0) = \frac{1}{2} \theta' V \theta + \frac{1}{\sqrt{n}} \theta' W_n + o_p(|\theta|^2) + o_p(1/n) \]

and then using the result in our Homework 1, Problem 5.

Re-express

\[ G_n(\theta) - G_n(0) = \underbrace{G(\theta) - G(0)}_{\text{deterministic component}} + \underbrace{[G_n(\theta) - G_n(0) - G(\theta) + G(0)]}_{\text{U-process}} \]

A Taylor approximation is used to show that the deterministic component \( G(\theta) - G(0) \) can be expressed as

\[ G(\theta) - G(0) = \frac{1}{2} \theta' V \theta + o(|\theta|^2). \]

The key is the second component (the random component, U-process). He shows that it can be expressed as

\[ G_n(\theta) - G_n(0) - G(\theta) + G(0) = \frac{1}{\sqrt{n}} \theta' W_n + o(|\theta|^2) + o_p(1/n) \]

uniformly over \( o_p(1) \) neighborhoods of \( \theta = 0 \), where \( W_n \) converges to a \( \mathcal{N}(0, \Delta) \) random vector.

A sketch of the details is as follows: For each \( \theta \in \Theta \) define

\[ f(z_1, z_2, \theta) = \mathbb{1}\{y_1 > y_2\} \left[ \mathbb{1}\{x'_1 \beta(\theta) > x'_2 \beta(\theta)\} - \mathbb{1}\{x'_1 \beta(0) > x'_2 \beta(0)\} \right] - G(\theta) + G(0). \]

Then

\[ G_n(\theta) - G_n(0) - G(\theta) + G(0) = U_{n,2} f(\cdot, \cdot, \theta). \]

Applying the Hoeffding decomposition, we can write

\[ U_{n,2} f(\cdot, \cdot, \theta) = \frac{1}{n} \sum_{i=1}^{n} \rho(Z_i, \theta) + \underbrace{U_{n,2} \nu(\cdot, \cdot, \theta)}_{\text{P-degenerate}} \]

where

\[ \rho(z, \theta) = f(z, P, \theta) + f(P, z, \theta) \]
where—recall our previously introduced notation—\( f(z, P, \theta) = E[f(z_1, z_2, \theta) | z_1 = z] \) and \( f(P, z, \theta) = E[f(z_1, z_2, \theta) | z_2 = z] \). The function \( \nu(\cdot, \cdot, \theta) \) is defined as
\[
\nu(z_1, z_2, \theta) = f(z_1, z_2, \theta) - f(z_1, P, \theta) - f(P, z_2, \theta)
\]

Using the definition of \( \tau(z, \theta) \) we have
\[
\rho(z, \theta) = \tau(z, \theta) - \tau(z, 0) - 2G'(\theta) + 2G(0)
\]

Using a Taylor approximation we have:
\[
\frac{1}{n} \sum_{i=1}^{n} \rho(Z_i, \theta) = \theta' W_n + o_p(\theta^2) + o_p(1/n)
\]

notice that \( G'(0) = 0 \) since \( G(\theta) \) is maximized at \( \theta = 0 \). By assumption, the remainder \( R_n(\theta) \) satisfies
\[
R_n(\theta) \leq C\theta^3 \frac{1}{n} \sum_{i=1}^{n} M(Z_i) = o_p(\theta^2) \quad (\text{last equality holds uniformly over } o_p(1) \text{ neighborhoods of } 0.)
\]

Therefore, uniformly over \( o_p(1) \) neighborhoods of \( \theta = 0 \),
\[
\frac{1}{n} \sum_{i=1}^{n} \rho(Z_i, \theta) = \frac{1}{\sqrt{n}} \theta' W_n + o_p(\theta^2), \quad \text{where } \quad W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \tau(Z_i, 0)}{\partial \theta} \xrightarrow{d} N(0, \Delta)
\]

The result we are after is:
\[
G_n(\theta) - G_n(0) = \frac{1}{2} \theta' V \theta + \frac{1}{\sqrt{n}} \theta' W_n + o_p(\theta^2) + o_p(1/n)
\]
in order to immediately apply Problem 5 in Homework 5. Therefore, we only have to show that
\[
U_{n,2} \nu(\cdot, \cdot, \theta) = o_p(1/n).
\]

This is where Lemma 2 comes into play. A strengthening of Lemma 2 is the following:

**Lemma 3** Suppose all the conditions of Lemma 2 hold and suppose that there exists \( \theta_0 \in \Theta \) such that \( f(\cdot, \theta_0) \equiv 0 \). If the parameterization is \( L^2(Q) \)-continuous at \( \theta_0 \), that is, if
\[
\int |f(\cdot, \theta)|^2 dQ \longrightarrow 0 \quad \text{as } \theta \longrightarrow \theta_0
\]
then
\[ n^{k/2} U_{n,k} f(\cdot, \theta) = o_p(1) \]
uniformly over \( o_p(1) \) neighborhoods of \( \theta_0 \).

The previous lemma yields the result \( U_{n,2} \nu(\cdot, \cdot, \theta) = o_p(1/n) \) immediately once we show that the class \( \{ \nu(\cdot, \cdot, \theta) : \theta \in \Theta \} \) is Euclidean:

Consider the class of functions \( \mathcal{H} = \{ h(\cdot, \cdot, \beta) : \beta \in \mathcal{B} \} \) where for each \((z_1, z_2) \in \mathcal{S}(Z) \times \mathcal{S}(Z)\) and each \( \beta \in \mathcal{B} \),
\[ h(z_1, z_2, \beta) = \mathbf{1}\{y_1 > y_2\} \mathbf{1}\{x_1' \beta > x_2' \beta\} \]

Then \( \mathcal{H} \) is Euclidean for constant envelope \( F = 1 \). To see this, define
\[ g(z_1, z_2, t; \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = \gamma_1 y_1 + \gamma_2 y_2 + \gamma_3 x_1 + \gamma_4' x_2 + \gamma_5 t \]
and the class of functions
\[ \mathcal{G} = \{ g(\cdot, \cdot, \cdot; \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) : \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5 \in \mathbb{R} \text{ and } \gamma_3, \gamma_4 \in \mathbb{R}^d \} \]

We use the definition of \( \mathcal{G} \) to show that the subgraphs of \( \mathcal{H} \) are a class of sets with polynomial discrimination:
\[ s(h(\cdot, \cdot, \beta)) = \{(z_1, z_2, t) \in \mathcal{S}(Z) \times \mathcal{S}(Z) \times \mathbb{R} : 0 < t < h(z_1, z_2, \beta)\} \]
\[ = \{(z_1, z_2, t) \in \mathcal{S}(Z) \times \mathcal{S}(Z) \times \mathbb{R} : \{y_1 - y_2 > 0\}, \{x_1' \beta - x_2' \beta > 0\}, \{t \geq 1\}^c, \{t > 0\}\} \]
\[ = \{(z_1, z_2, t) \in \mathcal{S}(Z) \times \mathcal{S}(Z) \times \mathbb{R} : \{g_1 > 0\}, \{g_2 > 0\}, \{g_3 \geq 1\}^c, \{g_4 > 0\}\} \]
which is the intersection of four sets, three of which belong to a polynomial class and the fourth is the complement of a set of polynomial class. Therefore \( s(h(\cdot, \cdot, \beta)) \) is of polynomial class and \( \mathcal{H} \) is Euclidean. Note trivially that the zero function is an element of \( \mathcal{H} \). By Lemma 3, we have \( U_{n,2} \nu(\cdot, \cdot, \theta) = o_p(1/n) \). Combined with the previous results this yields
\[ G_n(\theta) - G_n(0) = \frac{1}{2} \theta' V \theta + \frac{1}{\sqrt{n}} \theta' W_n + o_p(\|\theta\|^2) + o_p(1/n) \]
and by Problem 5 in Homework 1, we obtain
\[ \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V^{-1} \Delta V^{-1}), \]
where

\[ V = \frac{1}{2} E \left[ \frac{\partial^2 \tau(Z,0)}{\partial \theta \partial \theta'} \right], \quad \text{and} \quad \Delta = E \left[ \frac{\partial \tau(Z,0)}{\partial \theta} \frac{\partial \tau(Z,0)'}{\partial \theta} \right] \]