Lecture 4: U-Statistics & U-Process Minimizers

Bryan S. Graham, UC - Berkeley & NBER

September 4, 2015

Hoeffding (1948a) developed the basic theory of U-Statistics, a family of estimates which includes many familiar and interesting examples. This lecture reviews this theory. Standard references for the material presented here include Serfling (1980, Chapter 5), Lehmman (1999, Chapter 6) and van der Vaart (1998, Chapters 11 & 12).

The basic theory of U-Statistics allows for a presentation of large sample results for so-called U-Process minimizers. U-Process minimizers generalize the textbook M-Estimator (e.g., Wooldridge, 2010, Chapter 12). Honore and Powell (1994) is a basic reference on U-Process minimizers (see also Bose (2002)). Later I will will apply these results to a one-to-one matching model.

U-Statistics

Let $\{Z_i\}_{i=1}^{N}$ be a simple random sample. Let $h\left(Z_{i_1},\ldots,Z_{i_m}\right)$ be a symmetric *kernel* function. The assumption of symmetry is without loss of generality since we can always replace $h\left(Z_{i_1},\ldots,Z_{i_m}\right)$ with its average across permutations. A U-statistic is an average of the kernel $h\left(Z_{i_1},\ldots,Z_{i_m}\right)$ over all possible m-tuples of observations in the sample.

$$U_N = \binom{N}{m}^{-1} \sum_{\mathbf{i}\in C_{m,N}} h\left(Z_{i_1},\ldots,Z_{i_m}\right)$$

where $C_{m,N}$ denotes the set of all unique combinations of indices of size m drawn from the set $\{1,2,\ldots,N\}$.

The parameter of interest is

$$\theta = \mathbb{E}\left[h\left(Z_1,\ldots,Z_m\right)\right],$$

where the expectation is over m independent random draws from the target population.

Variance of U_N

For $s = 1, \ldots, m$ let

$$\bar{h}_s\left(z_1, \ldots, z_s\right) = \mathbb{E}\left[h\left(z_1, \ldots, z_s, Z_{s+1}, \ldots, Z_m\right)\right]$$

be the average over the last $m - s$ elements of $h(\cdot)$ holding the first s elements fixed. Note that since Z_{i_k} is independent of Z_{i_l} for all $k \neq l$ we have

$$\mathbb{E}\left[h\left(Z_1, \ldots, Z_s, Z_{s+1}, \ldots, Z_m\right) \middle| \left(Z_1, \ldots, Z_s\right) = \left(z_1, \ldots, z_s\right)\right] = \mathbb{E}\left[h\left(z_1, \ldots, z_s, Z_{s+1}, \ldots, Z_m\right)\right].$$

It is also useful to observe that

$$\mathbb{E}\left[\bar{h}_s\left(Z_1, \ldots, Z_s\right)\right] = \mathbb{E}\left[h\left(Z_1, \ldots, Z_m\right)\right] = \theta.$$

The variance of U_N has a special structure. Define, for $s = 1, \ldots, m$

$$\delta_s^2 = \mathbb{V}\left(\bar{h}_s\left(Z_1, \ldots, Z_s\right)\right).$$

The variance of U_N equals

$$\begin{aligned}
\mathbb{V}\left(U_N\right) &= \mathbb{V}\left(\binom{N}{m}^{-1} \sum_{\mathbf{i} \in C_{m,N}} h\left(Z_{i_1}, \ldots, Z_{i_m}\right)\right) \\
&= \binom{N}{m}^{-2} \sum_{\mathbf{i} \in C_{m,N}} \sum_{\mathbf{j} \in C_{m,N}} \mathbb{C}\left(h\left(Z_{i_1}, \ldots, Z_{i_m}\right), h\left(Z_{j_1}, \ldots, Z_{j_m}\right)\right). \qquad (1)
\end{aligned}$$

The form of the covariances in (1) depends on the number of indices in common. Let s be the number of indices in common in $Z_{i_1}, \ldots, Z_{i_m}$ and $Z_{j_1}, \ldots, Z_{j_m}$:

$$\begin{aligned}
\mathbb{C}\left(h\left(Z_{i_1}, \ldots, Z_{i_m}\right), h\left(Z_{j_1}, \ldots, Z_{j_m}\right)\right) &= \mathbb{E}\left[\left(h\left(Z_1, \ldots, Z_s, Z_{s+1}, \ldots, Z_m\right) - \theta\right)\right. \\
&\quad \left. \times \left(h\left(Z_1, \ldots, Z_s, Z'_{s+1}, \ldots, Z'_m\right) - \theta\right)\right] \qquad (2)
\end{aligned}$$

Conditional on $Z_1, \ldots, Z_s$ the two terms in (2) are independent so that, using the Law of Iterated Expectations,

$$\mathbb{E}\left[\left(\bar{h}_s\left(Z_1, \ldots, Z_s\right) - \theta\right)\left(\bar{h}_s\left(Z_1, \ldots, Z_s\right) - \theta\right)\right] = \delta_s^2.$$

Using the same argument yields

$$\mathbb{C}\left(\bar{h}_s\left(Z_1,\ldots,Z_s\right),h\left(Z_1,\ldots,Z_m\right)\right)=\delta_s^2.$$

By the Cauchy-Schwartz Inequality we have

$$\frac{\mathbb{C}\left(\bar{h}_s\left(Z_1,\ldots,Z_s\right),h\left(Z_1,\ldots,Z_m\right)\right)}{\delta_s\delta_m}\leq 1$$

and hence

$$\delta_s^2\leq\delta_m^2.$$

Continuing with this type of reasoning we get the weak ordering

$$\delta_1^2\leq\delta_2^2\leq\ldots\leq\delta_m^2.$$

In what follows we will assume that $\delta_m^2<\infty$.

To use these results to get an expression for $\mathbb{V}\left(U_N\right)$ begin by observing that the number of pairs of m-tuples $(i_1,\ldots,i_m)$ and $(j_1,\ldots,j_m)$ having exactly s elements in common is

$$\binom{N}{m}\binom{m}{s}\binom{N-m}{m-s}.$$

This follows since $\binom{N}{m}$ equals the number of ways of choosing $(i_1,\ldots,i_m)$ from the set $\{1,\ldots,N\}$. For each unique m-tuple there are $\begin{pmatrix}m\\s\end{pmatrix}$ ways of choosing a subset of size s from it. Having fixed the s indices in common there are then $\binom{N-m}{m-s}$ ways of choosing the $m-s$ non-common elements of $(j_1,\ldots,j_m)$ from the $N-m$ integers not already present in $(i_1,\ldots,i_m)$.

We therefore have

$$\begin{aligned}\mathbb{V}\left(U_N\right) &= \binom{N}{m}^{-2}\sum_{s=0}^{m}\binom{N}{m}\binom{m}{s}\binom{N-m}{m-s}\delta_s^2\\ &= \binom{N}{m}^{-1}\sum_{s=1}^{m}\binom{m}{s}\binom{N-m}{m-s}\delta_s^2.\\ &= \sum_{s=1}^{m}\frac{m!^2}{s!\left(m-s\right)!^2}\frac{\left(N-m\right)\left(N-m-1\right)\cdots\left(N-2m+s+1\right)}{N\left(N-1\right)\cdots\left(N-m+1\right)}\delta_s^2. \qquad (3)\end{aligned}$$

To understand this expression note that each of the covariances in (1) above have $s=0,\ldots,m$ elements in common. The coefficients on the δ_s^2 in (3) give the number of covariances with

s elements in common. Also note that $\delta_0^2 = 0$.

The coefficient on δ_1^2 is

$$\begin{aligned}\frac{m!^2}{1!\,(m-1)!^2}\frac{(N-m)\,(N-m-1)\cdots(N-2m+1+1)}{N\,(N-1)\cdots(N-m+1)} &= m^2\frac{\overbrace{(N-m)\,(N-m-1)\cdots(N-2m+2)}^{\text{m-1 terms}}}{\underbrace{N\,(N-1)\cdots(N-m+1)}_{\text{mterms}}} \\ &\simeq \frac{m^2}{N}.\end{aligned}$$

The coefficient on δ_2^2 is $O\left(N^{-2}\right)$ etc. We therefore have

$$\mathbb{V}\left(U_N\right) = \frac{m^2}{N}\delta_1^2 + O\left(N^{-2}\right)$$

and also that $\mathbb{V}\left(\sqrt{N}\left(U_N - \theta\right)\right) \rightarrow m^2\delta_1^2$ as $N \rightarrow \infty$.

If $\delta_1 = 0$ we say that U_N is a degenerate U-Statistic with degeneracy of order 1. I will not consider the properties of degenerate U-Statistics here, although this situation is more than a technical curiosity (cf., Graham, 2015).

First projection of U_N

The arguments outlined so far provide expressions for the mean and variance of U_N. To conduct inference we need an asymptotic normality result. While U_N is a sum of identically distributed random variables, not all elements of its summand are independent of one another. We cannot apply a standard central limit theorem (CLT).

To show asymptotic normality of $\sqrt{N}U_N$ we will therefore proceed as follows. First, we will construct a statistic U_N^* with the property that $\sqrt{N}U_N^*$ obeys a standard CLT. Second, we will show that $\sqrt{N}U_N$ converges in mean square to $\sqrt{N}U_N^*$. Since $\sqrt{N}U_N^*$ and $\sqrt{N}U_N$ are asymptotically equivalent, their limit distributions coincide. The basic strategy is to construct a statistic, the properties of which are easy to understand, and show that this statistic, in large enough samples, is approximately equal to the original statistic of interest, the properties of which were not directly apparent at the outset.

In order to use a CLT we'd like our approximation U_N^* to be a sum of independent and identically distributed random variables.

To simplify the argument assume that $m = 2$. To understand the how to construct U_N^* begin by considering the L^2 projection of U_N onto just the first observation Z_1:

$$\begin{aligned}\mathbb{E}\left[\left.U_N\right|Z_1\right] &= \binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{i<j}\mathbb{E}\left[\left.h\left(Z_i,Z_j\right)\right|Z_1\right] \\ &= \binom{N}{2}^{-1}(N-1)\bar{h}_1\left(Z_1\right)+\binom{N}{2}^{-1}\left(\binom{N}{2}-(N-1)\right)\theta \\ &= \frac{2}{N}\left\{\bar{h}_1\left(Z_1\right)-\theta\right\}+\theta. \qquad (4)\end{aligned}$$

The second equality follows because $\mathbb{E}\left[\left.h\left(Z_i,Z_j\right)\right|Z_1\right]=\bar{h}_1\left(Z_1\right)$ if either i or j equals 1 (which occurs $N-1$ times). In all other cases, by random sampling, $\mathbb{E}\left[\left.h\left(Z_i,Z_j\right)\right|Z_1\right]=\mathbb{E}\left[h\left(Z_i,Z_j\right)\right]=\theta$ (which occurs $\binom{N}{2}-(N-1)$ times). Second recall, that for $X_1,\ldots,X_K$ all independent, we have $\mathbb{E}\left[\left.Y\right|X_1,\ldots,X_K\right]=\sum_{k=1}^{K}\mathbb{E}\left[\left.Y\right|X_k\right]-(K-1)\mathbb{E}\left[Y\right]$. Using this fact and (4) yields

$$\begin{aligned}\mathbb{E}\left[\left.U_N\right|Z_1,\ldots,Z_N\right] &= \sum_{i=1}^{N}\mathbb{E}\left[\left.U_N\right|Z_i\right]-(N-1)\mathbb{E}\left[U_N\right] \\ &= \frac{2}{N}\sum_{i=1}^{N}\left\{\bar{h}_1\left(Z_i\right)-\theta\right\}+N\theta-(N-1)\mathbb{E}\left[U_N\right] \\ &= \frac{2}{N}\sum_{i=1}^{N}\left\{\bar{h}_1\left(Z_i\right)-\theta\right\}+\theta.\end{aligned}$$

Define the centered average

$$\tilde{h}_1\left(Z_i\right)=\bar{h}_1\left(Z_i\right)-\theta.$$

The projection of $U_N-\theta$ onto the set of all statistics of the form $\sum_{i=1}^{N}g\left(Z_i\right)$ is thus given by

$$U_N^*=\frac{m}{N}\sum_{i=1}^{N}\left\{\bar{h}_s\left(Z_i\right)-\theta\right\}=\frac{m}{N}\sum_{i=1}^{N}\tilde{h}_1\left(Z_i\right). \qquad (5)$$

Equation (5) is often called the Hajek projection; see van der Vaart (1998, Chapters 11-12) for more details.

Since U_N^* is a sum of i.i.d. random variables with $\mathbb{V}\left(\tilde{h}_s\left(Z_i\right)\right)=m^2\delta_1^2$, a CLT gives

$$\sqrt{N}U_N^*\overset{D}{\rightarrow}\mathcal{N}\left(0,m^2\delta_1^2\right).$$

We will now show (see also Theorem 12.3 in van der Vaart (2000)), that

$$N\mathbb{E}\left[\left(U_N^*-\left(U_N-\theta\right)\right)^2\right]\rightarrow 0 \qquad (6)$$

as $N \rightarrow \infty$. Equation (6) implies that $\sqrt{N}\left(U_N - \theta\right)$ converges in mean square to $\sqrt{N}U_N^*$. This latter statistic is normally distributed in large samples and hence so will be our centered U-Statistic.

We have

$$N\mathbb{E}\left[\left(U_N^* - \left(U_N - \theta\right)\right)^2\right] = N\mathbb{V}\left(U_N^*\right) - 2N\mathbb{C}\left(U_N^*, U_N\right) + N\mathbb{V}\left(U_N\right).$$

Evaluating the covariance component first yields

$$\begin{aligned} N\mathbb{C}\left(U_N^*, U_N\right) &= N\mathbb{C}\left(\frac{m}{N}\sum_{i=1}^{N}\tilde{h}_1\left(Z_i\right), \binom{N}{m}^{-1}\sum_{\mathbf{j}\in C_{m,N}} h\left(Z_{j_1}, \ldots, Z_{j_m}\right)\right) \\ &= \frac{m}{\binom{N}{m}}\sum_{i=1}^{N}\sum_{\mathbf{j}\in C_{m,N}}\mathbb{C}\left(\bar{h}_1\left(Z_i\right), h\left(Z_{j_1}, \ldots, Z_{j_m}\right)\right). \end{aligned}$$

This covariance is zero unless $i \in \{j_1, \ldots, j_m\}$ and equals δ_1^2 otherwise (by the calculations above). For a fixed i the number of m-tuples containing i is $\binom{N-1}{m-1}$ and since there are N such i we get therefore get

$$\begin{aligned} N\mathbb{C}\left(U_N^*, U_N\right) &= \frac{m}{\binom{N}{m}}N\tbinom{N-1}{m-1}\delta_1^2 \\ &= m^2\delta_1^2. \end{aligned}$$

Since

$$\frac{\binom{N-1}{m-1}}{\binom{N}{m}} = \frac{(N-1)!}{(m-1)!\left((N-1)-(m-1)\right)!}\frac{m!\left(N-m\right)!}{N!} = \frac{m}{N}.$$

This gives

$$\begin{aligned} N\mathbb{E}\left[\left(U_N^* - \left(U_N - \theta\right)\right)^2\right] &= \mathbb{V}\left(\sqrt{N}U_N\right) - m^2\delta_1^2 \\ &\rightarrow 0 \end{aligned}$$

as $N \rightarrow \infty$. Since $\sqrt{N}U_N^* - \sqrt{N}\left(U_N - \theta\right)$ converges in mean square to zero, they are asymptotically equivalent, and hence

$$\sqrt{N}\left(U_N - \theta\right) \xrightarrow{D} \mathcal{N}\left(0, m^2\delta_1^2\right)$$

as needed.

Kendall's Tau

There are many applications of U-Statistics to non-parametric testing problems. I will give just one example here. See Lehmann (1999) for an exposition of many classic examples.

Consider the statistics (e.g., Hoeffding, 1948b)

$$K_N = \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} \operatorname{sgn}\left\{ \left(X_i - X_j\right)\left(Y_i - Y_j\right) \right\}. \tag{7}$$

Let $Z = (X, Y)$, then (7) is a U-Statistic of order 2 with kernel $h(z_1, z_2) = \operatorname{sgn}\{(x_1 - x_2)(y_1 - y_2)\}$. To calculate the Hajek projection we evaluate

$$\begin{aligned}
\bar{h}_1(z) &= \mathbb{E}\left[\operatorname{sgn}\{(x - X)(y - Y)\}\right] \\
&= \Pr((x - X)(y - Y) > 0) - \Pr((x - X)(y - Y) < 0) \\
&= \Pr(X > x, Y > y \,\text{or}\, X < x, Y < y) - \\
&\quad \Pr(X > x, Y < y \,\text{or}\, X < x, Y > y) \\
&= 1 - 2F_{XY}(x, \infty) - 2F_{XY}(\infty, y) + 4F_{XY}(x, y) \\
&= (1 - 2F_X(x))(1 - 2F_Y(y)) + 4(F_{XY}(x, y) - F_X(x) F_Y(y)).
\end{aligned}$$

To verify the above calculations it is helpful to draw a figure. Under the null of independence of X and Y $F_{XY}(x, y) = F_X(x) F_Y(y)$ so that

$$\bar{h}_1(z) = (1 - 2F_X(x))(1 - 2F_Y(y)).$$

Note that $U = (1 - 2F_X(X))$ is a $\mathcal{U}[-1, 1]$ random variable, independent of $V = (1 - 2F_X(Y))$, which is also a $\mathcal{U}[-1, 1]$ random variable. Under the independence null we therefore have $\theta = \mathbb{E}[UV] = 0$.

The variance of $\bar{h}_1(Z)$ equals

$$\begin{aligned}\delta_1 &\overset{H_0}{=} \mathbb{V}\left(\bar{h}_1(Z)\right)\\ &= \mathbb{V}(UV)\\ &= \mathbb{E}\left(U^2V^2\right)\\ &= \left[\int_{-1}^{1} u^2\frac{1}{2}\mathrm{d}u\right]\left[\int_{-1}^{1} v^2\frac{1}{2}\mathrm{d}v\right]\\ &= \left(\frac{1}{3}\right)^2\\ &= \frac{1}{9}.\end{aligned}$$

Using these results, and the general large sample results above, we have that *under the maintained null of independence of* X and Y

$$\sqrt{N}K_N \overset{H_0}{\to} \mathcal{N}\left(0, \frac{4}{9}\right).$$

We may reject the null of independence if $\sqrt{\frac{9N}{4}}\left|K_N\right| > z_{\alpha/2}$.

U-Process Minimizers

Honoré and Powell (1994) study the large sample properties of U-Process minimizers (see also Bose (2002)). They are motivated by problems due to censoring.

Let $\{Z_i\}_{i=1}^N$ be a sample of i.i.d random variables and consider the estimator $\hat{\beta}$ which minimizes

$$L_N(\beta) = \binom{N}{m}^{-1}\sum_{\mathbf{i}\in C_{m,N}} l\left(Z_{i_1},\ldots,Z_{i_m};\beta\right).$$

A mean value expansion gives, after some manipulation

$$\sqrt{N}\left(\hat{\beta}-\beta_0\right) = -\Gamma_0^{-1}\sqrt{N}\left[\binom{N}{m}^{-1}\sum_{\mathbf{i}\in C_{m,N}} \nabla_\beta l\left(Z_{i_1},\ldots,Z_{i_m};\beta_0\right)\right] + o_p(1)$$

where $\underset{N\to\infty}{\text{plim}}\nabla_{\beta\beta}L_N\left(\hat{\beta}\right) = \Gamma_0$, assumed invertible. To make connections to the basic theory of U-Statistics outlined above define

$$h\left(Z_{i_1},\ldots,Z_{i_m};\beta\right)=\nabla_{\beta}l\left(Z_{i_1},\ldots,Z_{i_m};\beta\right)$$

and also

$$\tilde{h}_1\left(z_1;\beta\right)=\mathbb{E}\left[h\left(z_{1_1},Z_{i_2}\ldots,Z_{i_m};\beta\right)\right].$$

A CLT gives

$$\frac{m}{\sqrt{N}}\sum_{i=1}^{N}\tilde{h}_1\left(Z_i;\beta_0\right)\xrightarrow{D}\mathcal{N}\left(0,m^2\Omega_0\right).$$

with

$$\Omega_0=\mathbb{E}\left[\tilde{h}_1\left(Z_i;\beta_0\right)\tilde{h}_1\left(Z_i;\beta_0\right)'\right].$$

Define

$$U_N\left(\beta_0\right)=\binom{N}{m}^{-1}\sum_{\mathbf{i}\in C_{m,N}}\nabla_{\beta}l\left(Z_{i_1},\ldots,Z_{i_m};\beta_0\right),\ U_N^{*}\left(\beta_0\right)=\frac{m}{N}\sum_{i=1}^{N}\tilde{h}_1\left(Z_i;\beta_0\right).$$

By our discussion of U-Statistics given above we have

$$N\mathbb{E}\left[\left(U_N^{*}\left(\beta_0\right)-U_N\left(\beta_0\right)\right)^2\right]\rightarrow 0$$

as $N\rightarrow\infty$ and hence, applying a Slutsky Theorem,

$$\sqrt{N}\left(\hat{\beta}-\beta_0\right)\xrightarrow{D}\mathcal{N}\left(0,m^2\Gamma_0^{-1}\Omega_0\Gamma_0^{-1}\right).$$

A rigorous derivation of this result requires considerably more care, but the heuristic argument given here is a simple combination of textbook arguments associated with M-estimation (e.g., Wooldridge, 2001, Chapter 12) and those outlined for U-Statistics above.

To construct an estimate of the asymptotic variance of $\hat{\beta}$ we compute

$$\hat{\tilde{h}}_1\left(Z_i;\hat{\beta}\right)=\binom{N-1}{m-1}^{-1}\sum_{\mathbf{j}\in C_{m-1,N-1}}h\left(Z_i,Z_{j_2}\ldots,Z_{j_m};\beta_0\right),$$

and then calculate

$$\begin{aligned}\hat{\Omega} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\tilde{h}}_1\left(Z_i;\hat{\beta}\right)\hat{\tilde{h}}_1\left(Z_i;\hat{\beta}\right)' \\ \hat{\Gamma} &= \binom{N}{m}^{-1}\sum_{\mathbf{i}\in C_{m,N}}\nabla_{\beta\beta}l\left(Z_{i_1},\ldots,Z_{i_m};\hat{\beta}\right).\end{aligned}$$

Application: partially linear logit

Consider the binary choice model

$$Y_i = 1\left(X_i'\beta_0 + g\left(W_i\right) - U_i \geq 0\right),$$

with U_i logistic. Here we assume that W_i is discretely-valued, but perhaps with many support points. An estimator which replaces the unknown function $g\left(W_i\right)$ with a vector of dummy variables for each support point of W_i may have poor finite sample properties and/or be difficult to compute.

Let i and j be two independent random draws. Recalling results from binary choice with panel data analysis we have that

$$\begin{aligned}\Pr\left(Y_i = 0, Y_j = 1\ X_i, X_j, Y_i + Y_j = 1, W_i = W_j\right) &= \frac{\exp\left(X_j'\beta_0 + g\left(W_j\right)\right)}{\exp\left(X_j'\beta_0 + g\left(W_j\right)\right) + \exp\left(X_i'\beta_0 + g\left(W_j\right)\right)} \\ &= \frac{\exp\left(\left(X_j - X_i\right)'\beta_0\right)}{1 + \exp\left(\left(X_j - X_i\right)'\beta_0\right)}.\end{aligned}$$

If we let

$$S_{ij} = \operatorname{sgn}\left\{Y_j - Y_i\right\},$$

we may base estimation of β_0 on the U-Process

$$L_N\left(\beta\right) = \binom{N}{2}^{-1}\sum_{i=1}^{N}\sum_{j<i}\mathbf{1}\left(W_i = W_j\right)\left|S_{ij}\right|\left\{S_{ij}\left(X_j - X_i\right)'\beta - \ln\left[1 + \exp\left(S_{ij}\left(X_j - X_i\right)'\beta\right)\right]\right\}.$$

To construct an estimate of the asymptotic variance of $\hat{\beta}$ first define

$$\hat{\hat{h}}_1\left(Z_i;\hat{\beta}\right) = \frac{1}{N-1}\sum_{j=1, j\neq i}^{N}\mathbf{1}\left(Z_i = Z_j\right)\left|S_{ij}\right|\left\{\mathbf{1}\left(S_{ij} = 1\right) - \frac{\exp\left(\left(X_j - X_i\right)'\hat{\beta}\right)}{1 + \exp\left(\left(X_j - X_i\right)'\hat{\beta}\right)}\right\}\left(X_j - X_i\right),$$

and then compute

$$\begin{aligned}\hat{\Gamma} &= -\frac{2}{N\left(N-1\right)}\sum_{i=1}^{N}\sum_{j<i}\mathbf{1}\left(Z_i = Z_j\right)\left|S_{ij}\right|\left\{\frac{\exp\left(\left(X_j - X_i\right)'\hat{\beta}\right)}{\left[1 + \exp\left(\left(X_j - X_i\right)'\hat{\beta}\right)\right]^2}\right\}\left(X_j - X_i\right)\left(X_j - X_i\right)' \\ \hat{\Omega} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\hat{h}}_1\left(Z_i;\hat{\beta}\right)\hat{\hat{h}}_1\left(Z_i;\hat{\beta}\right)'.\end{aligned}$$

References

[1] Bose, Arup. (2002). "U statistics and M_m estimates," *Uncertainty and Optimality: Probability, Statistics and Operations Research*: 257 - 292 (J. C. Misra, Ed.). Singapore: World Scientific Publishing.

[2] Ferguson, Thomas S. (2005). "U-Statistics," *UCLA Lecture Note.*

[3] Graham, Bryan S. (2015). "An econometric model of link formation with degree heterogeneity," *CEMMAP Working Paper No. CWP43/15.*

[4] Hoeffding, Wassily. (1948a). "A Class of Statistics with asymptotically normal distribution," *Annals of Mathematical Statistics* 19 (3): 293 - 325.

[5] Hoeffding, Wassily. (1948b). "A non-parametric test of independence," *Annals of Mathematical Statistics* 19 (4): 546 - 557.

[6] Lehmann, E. L. (1999). *Elements of Large-Sample Theory.* New York: Springer.

[7] Powell, James L. and Bo E. Honore. (1994). "Pairwise difference estimators of censored and truncated regression models," *Journal of Econometrics* 64 (1-2): 241 - 278.

[8] Serfling, Robert J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley & Sons, Inc.

[9] van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

[10] Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.