U-Statistics

An Introduction to the Econometrics of Networks

CEMFI, November 28th to December 2nd, 2022

Bryan S. Graham

University of California - Berkeley

U-Statistics

Introduced by Wassily Hoeffding (1948).

Arise frequently in semiparametric econometrics.

Also useful for the analysis of dyadic data (and network data more generally).

References: Lehmann (1999, Chapter 6), van der Vaart (2000, Chapter 12) and Ferguson (2005).

A familiar example

Let

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i$$

be the sample mean and

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \bar{X}_N\right)^2$$

the sample variance.

It turns out that S^2 is a second order U-statistic.

Sample variance

$$
\begin{aligned}
S^2 &= \frac{1}{2N\left(N-1\right)} \sum_{i=1}^{N} \sum_{j=1}^{N} \left((X_i - \bar{X}_N)^2 + \left(X_j - \bar{X}_N\right)^2 \right) \\
&= \frac{1}{2N\left(N-1\right)} \sum_{i=1}^{N} \sum_{j=1}^{N} \left((X_i - \bar{X}_N) - \left(X_j - \bar{X}_N\right) \right)^2 \\
&= \frac{1}{N\left(N-1\right)} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{2} \left(X_i - X_j\right)^2 \\
&= \frac{2}{N\left(N-1\right)} \sum_{i=1}^{N-1} \sum_{j=1+1}^{N} \frac{1}{2} \left(X_i - X_j\right)^2 \\
&= \binom{N}{2}^{-1} \sum_{i<j} \frac{1}{2} \left(X_i - X_j\right)^2 .
\end{aligned}
$$

Unbiasedness under random sampling

If $\{X_i\}_{i=1}^{N}$ are i.i.d random draws from F_Z, then $\mathbb{E}\left[S^2\right]$ is unbiased for $\mathbb{V}\left(X_9\right)$:

$$\begin{aligned}
\mathbb{E}\left[S^2\right] &= \frac{1}{2}\mathbb{E}\left[(X_9 - X_{19})^2\right] \\
&= \frac{1}{2}\mathbb{E}\left[(X_9 - \mathbb{E}\left[X_9\right] - (X_{19} - \mathbb{E}\left[X_{19}\right]))^2\right] \\
&= \frac{1}{2}\mathbb{E}\left[(X_9 - \mathbb{E}\left[X_9\right])^2 + (X_{19} - \mathbb{E}\left[X_{19}\right])^2\right] \\
&= \mathbb{E}\left[(X_9 - \mathbb{E}\left[X_9\right])^2\right] \\
&= \mathbb{V}\left(X_9\right)
\end{aligned}$$

(X_9 and X_{19} are my favorite draws under random sampling when when $N \geq 19$).

Definition: U-Statistic

Let $\{X_i\}_{i=1}^N$ be a simple random sample from F_X.

Let $h\left(X_{i_1},\ldots,X_{i_m}\right)$ be a symmetric *kernel* function.

(We can always replace $h\left(X_{i_1},\ldots,X_{i_m}\right)$ with its average across permutations).

A U-statistic is an average of the kernel $h\left(X_{i_1},\ldots,X_{i_m}\right)$ over all possible m-tuples of observations in the sample.

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

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

Definition: U-Statistic

The parameter of interest is

$$\theta = \mathbb{E}\left[U_N\right] = \mathbb{E}\left[h\left(X_1, \ldots, X_m\right)\right],$$

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

Our goals for today:

1. U_N is unbiased for θ, what about $\mathbb{V}\left(U_N\right)$?

2. large sample theory / asymptotic normality;

3. extension to M-estimation type problems.

Kendall's Tau

$\{(X_i, Y_i)\}_{i=1}^N$ are i.i.d random draws from $F_{X,Y}$.

The probability that a pair of observations are concordant is

$$c = \text{Pr}\left(X_i > X_j \cap Y_i > Y_j\right) \cup \text{Pr}\left(X_i < X_j \cap Y_i < Y_j\right)$$

Kendall's Tau is the population proportion of concordant pairs minus the population proportion of discordant pairs

$$\begin{aligned} \tau &= c - (1 - c) \\ &= 2c - 1. \end{aligned}$$

A "nonparametric" measure of correlation/monotonicity: -1 (never concordant) and 1 (always concordant).

Kendall's Tau (continued)

Let $Z_i = (X_i, Y_i)$ and define the kernel $h\left(Z_i, Z_j\right)$ as

$h\left(Z_i, Z_j\right) = 2\left[\mathbf{1}\left(X_i > X_j\right)\mathbf{1}\left(Y_i > Y_j\right) + \mathbf{1}\left(X_i < X_j\right)\mathbf{1}\left(Y_i < Y_j\right)\right] - 1.$
With some work it is possible to show that

$$h\left(Z_i, Z_j\right) = \left(1 - 2\mathbf{1}\left(X_i < X_j\right)\right)\left(1 - 2\mathbf{1}\left(Y_i < Y_j\right)\right),$$

which will be useful later.

An unbiased estimate of τ is

$$\widehat{\tau} = \binom{N}{2}^{-1} \sum_{i<j} \frac{1}{2} h\left(Z_i, Z_j\right).$$

(we will symmetrize $h\left(Z_i, Z_j\right)$ later).

Variance

The variance of U_N is

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

(Some of) the summands in $\mathbb{V}(U_N)$ covary.

Fortunately this dependence is structured.

Variance (continued)

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

$$\bar{h}_s(x_1, \ldots, x_s) = \mathbb{E}\left[h\left(x_1, \ldots, x_s, X_{s+1}, \ldots, X_m\right)\right]$$

be the average over the last $m - s$ elements of $h(\cdot)$ holding the first s elements fixed.

Note that since X_{i_k} is independent of X_{i_l} for all $k \neq l$ we have

$$\begin{aligned} &\mathbb{E}\left[h\left(X_1, \ldots, X_s, X_{s+1}, \ldots, X_m\right) \middle| (X_1, \ldots, X_s) = (x_1, \ldots, x_s)\right] \\ &\qquad = \mathbb{E}\left[h\left(x_1, \ldots, x_s, X_{s+1}, \ldots, X_m\right)\right]. \end{aligned}$$

It is also useful to observe that

$$\mathbb{E}\left[\bar{h}_s(X_1, \ldots, X_s)\right] = \mathbb{E}\left[h(X_1, \ldots, X_m)\right] = \theta.$$

Variance (continued)

Define, for $s = 1, \ldots, m$,

$$\delta_s^2 = \mathbb{V}\left(\bar{h}_s\left(X_1, \ldots, X_s\right)\right).$$

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 $X_{i_1}, \ldots, X_{i_m}$ and $X_{j_1}, \ldots, X_{j_m}$:

$$\begin{aligned}
&\mathbb{C}\left(h\left(X_{i_1}, \ldots, X_{i_m}\right), h\left(X_{j_1}, \ldots, X_{j_m}\right)\right) \\
&\qquad = \mathbb{E}\Big[\left(h\left(X_1, \ldots, X_s, X_{s+1}, \ldots, X_m\right) - \theta\right) \\
&\qquad\qquad \times \left(h\left(X_1, \ldots, X_s, X'_{s+1}, \ldots, X'_m\right) - \theta\right)\Big] \quad (2)
\end{aligned}$$

Variance (continued)

Conditional on $X_1, \ldots, X_s$ the two terms in (2) are independent so that, using the Law of Iterated Expectations,

$$\mathbb{C}\left(h\left(X_{i_1}, \ldots, X_{i_m}\right), h\left(X_{j_1}, \ldots, X_{j_m}\right)\right)$$
$$= \mathbb{E}\left[\left(\bar{h}_s\left(X_1, \ldots, X_s\right) - \theta\right)\left(\bar{h}_s\left(X_1, \ldots, X_s\right) - \theta\right)\right]$$
$$= \delta_s^2.$$

Using the same argument yields

$$\mathbb{C}\left(\bar{h}_s\left(X_1, \ldots, X_s\right), h\left(X_1, \ldots, X_m\right)\right) = \delta_s^2.$$

Variance (continued)

By the Cauchy-Schwartz Inequality we have

$$\frac{\mathbb{C}\left(\bar{h}_s\left(X_1,\ldots,X_s\right),h\left(X_1,\ldots,X_m\right)\right)}{\delta_s\delta_m}\leq 1$$

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 $\delta_m^2 < \infty$.

Variance (continued)

To use these results to get an expression for $\mathbb{V}(U_N)$ 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

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

Variance (continued)

This follows since:

1. $\binom{N}{m}$ equals the number of ways of choosing $(i_1, \ldots, i_m)$ from the set $\{1, \ldots, N\}$.

2. For each unique m-tuple there are $\binom{m}{s}$ ways of choosing a subset of size s from it.

3. 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 $(i_1, \ldots, i_m)$.

Variance (continued)

We therefore have

$$\begin{aligned}
\mathbb{V}\left(U_N\right) &= \binom{N}{m}^{-2} \sum_{s=0}^{m} \binom{N}{m}\binom{m}{s}\binom{N-m}{m-s}\delta_s^2 \\
&= \sum_{s=1}^{m} \left[\frac{m!^2}{s!\,(m-s)!^2} \times \right. && (3) \\
&\quad \left. \frac{(N-m)\,(N-m-1)\cdots(N-2m+s+1)}{N\,(N-1)\cdots(N-m+1)}\right] \delta_s^2 && (4)
\end{aligned}$$

Variance (continued)

To understand this expression note that each of the covariances in (4) above have $s = 0, \ldots, m$ elements in common.

The coefficients on the δ_s^2 in (4) give the number of covariances with s elements in common.

Also note that $\delta_0^2 = 0$.

Variance (continued)

The coefficient on δ_1^2 is

$$\frac{m!^2}{1!\,(m-1)!^2}\frac{(N-m)\,(N-m-1)\cdots(N-2m+1+1)}{N\,(N-1)\cdots(N-m+1)}$$

$$= m^2\frac{\overbrace{(N-m)\,(N-m-1)\cdots(N-2m+2)}^{\text{m-1 terms}}}{\underbrace{N\,(N-1)\cdots(N-m+1)}_{\text{mterms}}}$$

$$\simeq \frac{m^2}{N}.$$

Variance (continued)

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

$$\boxed{\mathbb{E}\left[U_N\right] = \theta, \quad \mathbb{V}\left(U_N\right) = \frac{m^2}{N}\delta_1^2 + O\left(N^{-2}\right)}$$

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

If $\delta_1 = 0$ we say that U_N is a degenerate U-Statistic with degeneracy of order 1.

Large sample theory

Basic idea:

1. We are interested in the asymptotic distribution of U_N (a priori complicated).

2. Find another statistics U_N^* with well-understood asymptotic distribution.

3. Show that U_N is "close enough" to U_N^* as $N \to \infty$ such that they have the same asymptotic distribution

Hajek Projection

The asymptotic properties of sums of independent random variables, appropriately scaled, are especially well-understood.

Let $X_1, X_2, \ldots, X_N$ be independent $K \times 1$ random vectors. Let $\mathcal{L}$ be the linear subspace containing of all functions of the form

$$\sum_{i=1}^{N} g_i \left(X_i\right) \tag{5}$$

for $g_i : \mathbb{R}^K \to \mathbb{R}$ arbitrary with $\mathbb{E}\left[g_i \left(X_i\right)^2\right] < \infty$ for $i = 1, \ldots, N$.

Hajek Projection (continued)

Next let Y be an arbitrary random variable with finite variance, but unknown distribution.

Use the Projection Theorem to approximate the statistic Y with one composed of a sum of independent random functions.

Such a sum, by appeal to a CLT, may be well-described by a normal distribution.

If the projection is also a very good approximation of Y, then the hope is that Y may be well-described by a normal distribution as well.

Hajek Projection (continued)

The projection of Y onto $\mathcal{L}$, equals

$$\Pi\left(Y|\,\mathcal{L}\right) = \sum_{i=1}^{N} \mathbb{E}\left[Y|\,X_i\right] - (N-1)\,\mathbb{E}\left[Y\right]. \tag{6}$$

To verify (6) it suffices to check the necessary and sufficient orthogonality condition of the Projection Theorem.

Hajek Projection: verification

It is helpful to observe that, for $j \neq i$,

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[Y \mid X_i\right] \mid X_j\right] &= \mathbb{E}\left[\mathbb{E}\left[Y \mid X_i\right]\right] \\ &= \mathbb{E}\left[Y\right], \qquad (7)\end{aligned}$$

due to independence of X_i and X_j and the law of iterated expectations.

In contrast, if $j = i$, then

$$\mathbb{E}\left[\mathbb{E}\left[Y \mid X_i\right] \mid X_i\right] = \mathbb{E}\left[Y \mid X_i\right]. \qquad (8)$$

Hajek Projection: verification (continued)

The orthogonality condition to verify, for $U = Y - \Pi\left(Y \mid \mathcal{L}\right)$, is

$$\begin{aligned} 0 =& \mathbb{E}\left[U\left(\sum_{j=1}^{N} g_j\left(X_j\right)\right)\right] \\ =& \sum_{j=1}^{N} \mathbb{E}\left[\mathbb{E}\left[U \mid X_j\right] g_j\left(X_j\right)\right] \end{aligned}$$

Next observe that, using (7) and (8),

$$\begin{aligned} \mathbb{E}\left[U \mid X_j\right] &= \mathbb{E}\left[Y \mid X_j\right] - \sum_{i=1}^{N} \mathbb{E}\left[\mathbb{E}\left[Y \mid X_i\right] \mid X_j\right] + (N-1)\,\mathbb{E}\left[Y\right] \\ &= \mathbb{E}\left[Y \mid X_j\right] - \mathbb{E}\left[Y \mid X_j\right] - (N-1)\,\mathbb{E}\left[Y\right] + (N-1)\,\mathbb{E}\left[Y\right] \\ &= 0, \end{aligned}$$

for $j = 1, \ldots, N$.

Hajek Projection: iid simplification

If, in addition to independence, we have that (i) $\{X_i\}_{i=1}^N$ are identically distributed and (ii) $Y = h(X_1, \ldots, X_N)$ is a permutation symmetric function of $\{X_i\}_{i=1}^N$, then

$$\begin{aligned}\mathbb{E}[Y| X_i = x] &= \mathbb{E}[Y| X_1 = x] \\ &= \mathbb{E}[h(x, X_2 \ldots, X_N)] \\ &\overset{def}{\equiv} \bar{h}_1(x)\end{aligned}$$

for all $i = 1, \ldots, N$. Since $h_1(x)$ does not depend on i it follows that (6) simplifies, in this case, to

$$\boxed{\Pi(Y| \mathcal{L}) = \sum_{i=1}^{N} \bar{h}_1(X_i) - (N-1)\mathbb{E}[Y]} \qquad (9)$$

Large Sample Theory

Let $\{Y_N\}$ be a sequence of statistics indexed by the sample size and $\mathcal{L}_N$ a corresponding sequence of linear subspaces of form (5).

Goal: use the limiting distribution of $\sqrt{N}\left(\Pi\left(Y_N|\,\mathcal{L}_N\right) - \Pi\left(Y_N|\,1\right)\right)$ to approximate that of $\sqrt{N}\left(Y_N - \Pi\left(Y_N|\,1\right)\right)$.

Valid if these two statistics converge in mean square (to one another).

Attractive because in many cases of interest the asymptotic sampling distribution of $\sqrt{N}\left(\Pi\left(Y_N|\,\mathcal{L}_N\right) - \Pi\left(Y_N|\,1\right)\right)$ is straightforward to derive, whereas that of $\sqrt{N}\left(Y_N - \Pi\left(Y_N|\,1\right)\right)$ may be ex ante non-obvious.

Large Sample Theory (continued)

The "Analysis of Variance" decomposition for projections gives

$$\begin{aligned}\|Y_N - \Pi(Y_N|\,1)\|^2 = {} & \|Y_N - \Pi(Y_N|\,\mathcal{L}_N)\|^2 \\ & + \|\Pi(Y_N|\,\mathcal{L}_N) - \Pi(Y_N|\,1)\|^2,\end{aligned}$$

which, after some re-arrangement, yields

$$\begin{aligned}N\|Y_N - \Pi(Y_N|\,\mathcal{L}_N)\|^2 = {} & N\|Y_N - \Pi(Y_N|\,1)\|^2 \\ & - N\|\Pi(Y_N|\,\mathcal{L}_N) - \Pi(Y_N|\,1)\|^2.\end{aligned}$$

Or, invoking the covariance inner product, that $\Pi(Y|\,1) = \mathbb{E}[Y]$, as well as the definition of variance

$$N\mathbb{E}\left[(Y_N - \Pi(Y_N|\,\mathcal{L}_N))^2\right] = N\mathbb{V}(Y_N) - N\mathbb{V}(\Pi(Y_N|\,\mathcal{L}_N)). \quad (10)$$

Large Sample Theory

If the limits of $N\mathbb{V}(Y_N)$ and $N\mathbb{V}(\Pi(Y_N|\,\mathcal{L}_N))$ coincide as $N \to 0$ we have that $\sqrt{N}(Y_N - \Pi(Y_N|\,\mathcal{L}_N))$ converges in mean square to zero.

This means that $\sqrt{N}Y_N$ and $\sqrt{N}\Pi(Y_N|\,\mathcal{L}_N)$ will have identical limit distributions.

Application to U-Statistics

The Hajek Projection of U_N onto $\mathcal{L}$ equals, from (6) above,

$$\Pi\left(U_N \middle| \mathcal{L}_N\right) = \sum_{i=1}^{N} \mathbb{E}\left[U_N \middle| X_i\right] - (N-1)\,\mathbb{E}\left[U_N\right]. \tag{11}$$

To simplify the argument assume that $m = 2$. The L^2 projection of U_N onto just the first observation X_1 is

$$\begin{aligned}
\mathbb{E}\left[U_N \middle| X_1\right] &= \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathbb{E}\left[h\left(X_i, X_j\right) \middle| X_1\right] \\
&= \binom{N}{2}^{-1} (N-1)\,\bar{h}_1\left(X_1\right) + \binom{N}{2}^{-1} \left(\binom{N}{2} - (N-1)\right)\theta \\
&= \frac{2}{N}\left\{\bar{h}_1\left(X_1\right) - \theta\right\} + \theta. \qquad (12)
\end{aligned}$$

Application to U-Statistics (continued)

The second equality follows because $\mathbb{E}\left[h\left(X_i, X_j\right)\middle|\, X_1\right] = \bar{h}_1\left(X_1\right)$ if either i or j equals 1 (which occurs $N-1$ times).

In all other cases, by random sampling, $\mathbb{E}\left[h\left(X_i, X_j\right)\middle|\, X_1\right] = \mathbb{E}\left[h\left(X_i, X_j\right)\right]$ θ (which occurs $\binom{N}{2} - (N-1)$ times).

Substituting (12) into (11) yields

$$\Pi\left(U_N - \theta\middle|\, \mathcal{L}_N\right) = \frac{2}{N}\sum_{i=1}^{N}\left\{\bar{h}_1\left(X_i\right) - \theta\right\}.$$

For the general $m \geq 2$ case a similar calculation gives

$$\Pi\left(U_N - \theta\middle|\, \mathcal{L}_N\right) = \frac{m}{N}\sum_{i=1}^{N}\left\{\bar{h}_1\left(X_i\right) - \theta\right\}.$$

Application to U-Statistics (continued)

Since $\Pi\left(U_N - \theta \middle| \mathcal{L}_N\right)$ is a sum of i.i.d. random variables with $\mathbb{V}\left(\bar{h}_1\left(X_1\right) - \theta\right) = \delta_1^2$, a CLT gives

$$\sqrt{N}\Pi\left(U_N - \theta \middle| \mathcal{L}_N\right) \xrightarrow{D} \mathcal{N}\left(0, m^2\delta_1^2\right).$$

Our (combinatoric) variance calculations gave

$$\mathbb{V}\left(\sqrt{N}\left(U_N - \theta\right)\right) \to m^2\delta_1^2$$

as $N \to \infty$.

Application to U-Statistics (continued)

Therefore

$$N\mathbb{V}\left(U_N\right) - N\mathbb{V}\left(\Pi\left(U_N \mid \mathcal{L}_N\right)\right) \rightarrow 0$$

as $N \rightarrow \infty$, in turn implying that $\sqrt{N}\left(U_N - \theta\right)$ convergences in mean square to $\sqrt{N}\Pi\left(U_N - \theta \mid \mathcal{L}_N\right)$ and hence that

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

as needed.

Limit Theory for Kendall's Tau

First we symmetrize the kernel:

$$
\begin{aligned}
h\left(Z_i, Z_j\right) =& \frac{1}{2}\Big[\left(1 - 2\mathbb{1}\left(X_i < X_j\right)\right)\left(1 - 2\mathbb{1}\left(Y_i < Y_j\right)\right) \\
&+ \left(1 - 2\mathbb{1}\left(X_i < X_j\right)\right)\left(1 - 2\mathbb{1}\left(Y_i < Y_j\right)\right)\Big].
\end{aligned}
$$

Limit Theory for Kendall's Tau (continued)

To calculate the projection we evaluate

$$\begin{aligned}\mathbb{E}\left[h\left(Z_1,Z_2\right)\middle|\,Z_1\right] =&\frac{1}{2}\left(1-2\left[1-F_X\left(X_1\right)\right]\right)\left(1-2\left[1-F_Y\left(Y_1\right)\right]\right)\\ &+\frac{1}{2}\left(1-2F_X\left(X_1\right)\right)\left(1-2F_Y\left(Y_1\right)\right)\\ =&\frac{1}{2}\left(2F_X\left(X_1\right)-1\right)\left(2F_Y\left(Y_1\right)-1\right)\\ &+\frac{1}{2}\left(1-2F_X\left(X_1\right)\right)\left(1-2F_Y\left(Y_1\right)\right).\end{aligned}$$

Limit Theory for Kendall's Tau (continued)

Next observe that $U_1 \overset{d}{=} 1 - 2F_X(X_1)$ and $V_1 \overset{d}{=} 1 - 2F_Y(Y_1)$ with U_1 and V_1 uniform on $[-1, 1]$.

We therefore have

$$\mathbb{E}\left[h(Z_1, Z_2)|\, Z_1\right] = \bar{h}_1(Z_1) \overset{d}{=} U_1 V_1.$$

Under the null of independence of X_1 and Y_1

$$\delta_1^2 = \mathbb{V}(U_1 V_1) \overset{H_0}{=} \frac{1}{9}$$

(If $U \sim$ Uniform $[a, b]$, then $\mathbb{E}[U] = (a+b)/2$ and $\mathbb{V}[U] = (b-2)^2/12$).

Limit Theory for Kendall's Tau (continued)

Putting things together we get

$$\boxed{\sqrt{N}\,(\hat{\tau}-\tau)\ \underset{H_0}{\overset{D}{\to}}\ \mathcal{N}\left(0,\frac{4}{9}\right).}$$

No need to calculate a variance to test the null of no-concordance.

Exercises: Derive limit theory for "general" case.

U-Process Minimizers

Honoré & Powell (1994) study the large sample properties of U-Process minimizers.

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

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

where $\operatorname*{plim}_{N\to\infty} \nabla_{\beta\beta} L_N\left(\widehat{\beta}\right) = \Gamma_0$, assumed invertible.

U-Process Minimizers (continued)

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

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

U-Process Minimizers (continued)

Define

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

$$U_N^*(\beta_0) = \frac{m}{N} \sum_{i=1}^{N} \tilde{h}_1(Z_i; \beta_0).$$

By our discussion of U-Statistics given above we have

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

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

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

U-Process Minimizers (continued)

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

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

and then calculate

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

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.

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.

Application: partially linear logit (continued)

Let i and j be two independent random draws.

Recalling results from binary choice with panel data analysis we have that

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

Application: partially linear logit (continued)

If we let

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

we may base estimation of β_0 on the U-Process

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

Application: partially linear logit (continued)

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

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

Application: partially linear logit (continued)

We can then compute

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

Wrapping Up

"Modern" applications may involve kernels which are indexed by N (e.g., semiparametric M-Estimation, Sparse Network Asymptotics).

Degeneracy: what happens if $\delta_1^2 = 0$? See Menzel (2021).

Two-sample U-Statistics (useful in Bi-Partite settings).

Exercises:

See Ferguson (2005) for many classic examples from nonparametric statistics which you can use as practice problems.

Compare exact distribution of sample variance in Gaussian model with U-Statistic large sample theory.

*References

Ferguson, T. S. (2005). U-statistics. University of California - Los Angeles.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19(3), 293 − 325.

Honoré, B. E. & Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, 64(1-2), 241 − 278.

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

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica*, 89(5), 2143 – 2188.

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