MAXIMAL INEQUALITIES FOR DEGENERATE U-PROCESSES WITH APPLICATIONS TO OPTIMIZATION ESTIMATORS [1]

Running Head: Maximal Inequalities for Degenerate U-processes.

By Robert P. Sherman

Bellcore

Abstract

Maximal Inequalities for degenerate U-processes of order k, $k \geq 1$, are established. The results rest on a moment inequality (due to Bonami (1970)) for kth-order forms, and extensions of chaining and symmetrization inequalities from the theory of empirical processes. Rates of uniform convergence are obtained.

The maximal inequalities can be used to determine the limiting distribution of estimators that optimize criterion functions having U-process structure. As an application, a semiparametric regression estimator that maximizes a U-process of order three is shown to be $\sqrt{n}$-consistent and asymptotically normally distributed.

1. Introduction. Let $Z_1, \ldots, Z_n$ be independent observations from a distribution P on a set $\mathcal{S}$. Let k be a positive integer and $\mathcal{F}$ a class of real-valued functions on $\mathcal{S}^k = \mathcal{S} \otimes \cdots \otimes \mathcal{S}$ (k factors). For each $f \in \mathcal{F}$, define

$$U_n^k f = (n)_k^{-1} \sum_{\boldsymbol{i_k}} f(Z_{i_1}, \ldots, Z_{i_k})$$

where $(n)_k = n(n-1)\cdots(n-k+1)$ and $\boldsymbol{i_k} = (i_1, \ldots, i_k)$ ranges over the $(n)_k$ ordered k-tuples of distinct integers from the set $\{1, \ldots, n\}$. By analogy with the empirical measure P_n that places mass n^{-1} at each Z_i, U_n^k can be viewed as a random probability measure putting mass $(n)_k^{-1}$ at each ordered k-tuple $(Z_{i_1}, \ldots, Z_{i_k})$. Note that $P_n \equiv U_n^1$. The function f need not be symmetric in its arguments. Apart from this, $U_n^k f$ is a U-statistic of order k in the sense of Serfling (1980, Chapter 5). The collection $\{U_n^k f : f \in \mathcal{F}\}$ is called a U-process of order k, and is said to be indexed by $\mathcal{F}$.

If, for each $f \in \mathcal{F}$,

$$Pf(s_1, \ldots, s_{i-1}, \cdot, s_{i+1}, \ldots, s_k) \equiv 0 \qquad i = 1, \ldots, k$$

[1] *MSC* 1991 *subject classifications.* Primary 62E20; secondary 60E15, 60G20, 60G99.

Key words and phrases. Maximal inequality, degenerate U-processes, empirical processes, chaining, symmetrization, polynomial classes of sets, Euclidean classes of functions, optimization estimator, semi-parametric estimation, generalized regression model.

then $\mathcal{F}$ is called a P-degenerate class of functions on $\mathcal{S}^k$, $U_n^k f$ is called a degenerate U-statistic of order k, and $\{U_n^k f : f \in \mathcal{F}\}$ is called a degenerate U-process of order k.

Let p be a positive integer. In this paper, we establish pth-moment maximal inequalities for degenerate U-processes indexed by classes of square-integrable functions. That is, for each $p \geq 1$ and $k \geq 1$, we obtain a bound for the moment

$$\mathbb{P} \sup_{\mathcal{F}} |n^{k/2} U_n^k f|^p$$

where $\mathbb{P}$ is the probability measure on the space on which the random elements $Z_1, \ldots, Z_n$ are defined, $\{U_n^k f : f \in \mathcal{F}\}$ is a degenerate U-process of order k, and the class of functions $\mathcal{F}$ comes equipped with an $\mathcal{L}^2$ pseudometric. We also state conditions under which these bounds are finite, uniformly in n.

Notice that a degenerate U-process of order one is a zero-mean empirical process. Pisier (1983) established a first-moment bound for such processes. Pollard (1990, Section 7) generalized Pisier's methods using Orlicz norms to obtain a bound for the moment

$$\mathbb{P}\Phi(\sup_{\mathcal{F}} |\sqrt{n} P_n f|)$$

where $\Phi(x) = \kappa_1 \exp(\kappa_2 x^2)$ for positive constants κ_1 and κ_2. Nolan and Pollard (1987) established moment inequalities for degenerate U-processes of order two. Implicit in their results is a bound for the moment

$$\mathbb{P}\Phi(\sup_{\mathcal{F}} |n U_n^2 f|)$$

where $\Phi(x) = \kappa_1 \exp(\kappa_2 |x|)$ for positive constants κ_1 and κ_2.

The bounds obtained by Pollard for the case $k = 1$, and those obtained by Nolan and Pollard for $k = 2$ are sharper than the corresponding bounds established in this paper. However, the methods employed by these authors do not generalize to cover degenerate U-processes of orders greater than two. This paper provides a method flexible enough to cover degenerate U-processes of arbitrary order.

Independently and almost simultaneously with this paper, Arcones and Giné (1991) established maximal inequalities for degenerate U-processes of arbitrary order. Implicit in their results is a bound for the moment

$$\mathbb{P}\Phi(\sup_{\mathcal{F}} |n^{k/2} U_n^k f|)$$

where $\Phi(x) = \kappa_1 \exp(\kappa_2 |x|^{2/k})$ for positive constants κ_1 and κ_2. These exponential moment bounds provide a beautiful generalization of the bounds mentioned above for the cases $k = 1$ and $k = 2$. The techniques used by these authors are similar, but distinct, from those employed in this paper. For example, they use a decoupling result for U-statistics established by de la Peña (1990) that plays a similar role to the symmetrization inequality established in Section 3.

Applications in the field of semiparametric estimation provided the original impetus for developing maximal inequalities for degenerate U-processes of orders greater than two. Cavanagh (1990) proposed a semiparametric rank estimator of the regression coefficients in a generalized regression model. His estimator maximizes a criterion function that is a U-process of order three. In Section 7, we show that his estimator is $\sqrt{n}$-consistent and asymptotically normally distributed. Degenerate U-processes enter the analysis through a simple decomposition of a U-process into a sum of degenerate U-processes. The uniformity result needed to establish the limiting distribution of the estimator requires the application of a uniform bound for degenerate U-processes of order three. This bound is established in Section 6, and is based on a maximal inequality for degenerate U-processes of order three proved in Section 5.

Ichimura (1988) proposed a semiparametric estimator of the regression coefficients in another generalized regression model. His estimator minimizes a criterion function that can be represented as a linear combination of U-processes of various orders (including those of orders three and higher) plus a process that has no effect on the limiting distribution of the estimator. Sherman (1991a) proved $\sqrt{n}$-consistency and asymptotic normality of Ichimura's estimator using the maximal inequalities established in this paper. Klein and Spady (1992) proposed a semiparametric estimator of the regression coefficients in a binary choice regression model. Their estimator maximizes a criterion function that can be represented as a linear combination of U-processes of various orders (again, including those of orders three and higher) plus a term which is negligible, asymptotically. As with Ichimura's estimator, the asymptotic distribution of the estimator of Klein and Spady can be determined using the maximal inequalities developed in this paper.

Applications are not limited to semiparametric regression estimators. Liu's (1990) generalized sample median, for example, maximizes a U-process of order k, $k \geq 2$. In future work, we hope to establish $\sqrt{n}$-consistency and asymptotic normality of her estimator, using the results in this paper.

In the next section, we establish a pth-moment maximal inequality for a general stochastic process. This is done by extending a chaining argument, due to Pisier (1983), to cover all integer moments. The next two sections provide the tools needed to specialize these results to degenerate U-processes. In Section 3, a symmetrization inequality commonly used in proving maximal inequalities for zero-mean empirical processes is generalized, and in Section 4, a moment inequality for kth-order forms is presented. The main results in the paper, namely, the pth-moment maximal inequalities for degenerate U-processes, are established in Section 5. In Section 6, we derive some useful consequences from the maximal inequalities, including rates of uniform convergence. Finally, in Section 7, we apply these results to determine the asymptotic distribution of Cavanagh's rank estimator.

2. Chaining Inequality. Consider a stochastic process $\{Z(t) : t \in T\}$, where T is an index set equipped with a pseudometric, d.

DEFINITION 1. For each $\epsilon > 0$, define the *packing number* $D(\epsilon, d, T)$ to be the largest number D for which there exist points $m_1, \ldots, m_D$ in T such that

$$d(m_i, m_j) > \epsilon \qquad \text{for} \quad i \neq j\,.$$

The packing numbers for the pseudometric space (d, T) tell us how big T is with respect to d, and will appear in the upper bound of the maximal inequality established below. For convenience, we will express the result in terms of an $\mathcal{L}^p$ norm.

CHAINING INEQUALITY. *Let Ψ be a convex, strictly increasing function on $[0, \infty)$, with $0 \leq \Psi(0) < 1$. Let p be a positive integer and write $\Psi_p(x)$ for $\Psi(x^p)$. Suppose the stochastic process $\{Z(t) : t \in T\}$ satisfies*

(i) *if $d(s,t) = 0$ then $Z(s) = Z(t)$ almost surely;*

(ii) *if $d(s,t) > 0$ then $I\!P\Psi_p(|Z(s) - Z(t)|/d(s,t)) \leq 1$;*

(iii) *there exists a point t_0 in T for which $\delta = \sup_T d(t, t_0) < \infty$;*

(iv) *the sample paths of Z are continuous.*

Then

$$\left\| \sup_T |Z(t) - Z(t_0)| \right\|_p \leq 2 \int_0^\delta \Psi_p^{-1}(D(x))\,dx$$

where $D(x)$ is short for the packing number $D(x, d, T)$.

PROOF. The proof rests on a simple convexity result. Suppose $Z_1, \ldots, Z_D$ are random variables and Δ is a positive real number such that $I\!P\Psi_p(|Z_i|/\Delta) \leq 1$ for each i. Then

$$\left\| \max_i |Z_i| \right\|_p \leq \Delta \Psi_p^{-1}(D)\,. \tag{1}$$

This follows from Jensen's inequality:

$$\Psi I\!P\left(\max_i |Z_i|^p/\Delta^p\right) \leq \sum_i I\!P\Psi(|Z_i|^p/\Delta^p) \leq D\,.$$

The bound will be applied to the increments of Z.

Define $\delta_i = \delta/2^i$ for $i = 0, 1, 2, \ldots$. Construct a sequence of maximal subsets $T(0), T(1), \ldots$ where $T(0) = \{t_0\}$, and for each i,

$$d(s,t) > \delta_i \quad \text{if} \quad s, t \in T(i) \quad \text{and} \quad s \neq t\,.$$

By definition of maximality, there exists a map γ_i from T into $T(i)$ for which $d(t, \gamma_i t) \le \delta_i$. Note that $T(i)$ contains at most $D(\delta_i)$ points.

Approximate $\sup_T |Z(t) - Z(t_0)|$ by $\max_{T(m)} |Z(t) - Z(t_0)|$ for some positive integer m. For each t in $T(m)$, define a chain of points leading from t to t_0:

$$t_m = t\,, \quad t_{m-1} = \gamma_{m-1} t_m\,, \quad \ldots\,, \quad t_1 = \gamma_1 t_2\,, \quad t_0 = \gamma_0 t_1\,.$$

By the triangle inequality followed by a crude bound,

$$\max_{T(m)} |Z(t) - Z(t_0)| \le \sum_{i=1}^{m} \max_{T(i)} |Z(t_i) - Z(t_{i-1})|\,.$$

Take $\mathcal{L}^p$ norms of both sides and apply the triangle inequality once again. Then apply (1) and the fact that $\delta_i = 2(\delta_i - \delta_{i+1})$ to get

$$\begin{aligned} \left\| \max_{T(m)} |Z(t) - Z(t_0)| \right\|_p &\le \sum_{i=1}^{m} \left\| \max_{T(i)} |Z(t_i) - Z(t_{i-1})| \right\|_p \\ &\le 2 \sum_{i=1}^{\infty} (\delta_i - \delta_{i+1}) \Psi_p^{-1}(D(\delta_i)) \\ &\le 2 \int_0^{\delta} \Psi_p^{-1}(D(x))\, dx\,. \end{aligned}$$

Let m tend to infinity, then appeal to monotone convergence and continuity of the sample paths to complete the proof. □

The basic chaining argument in the proof is due to Pisier (1983). Refinements due to Nolan and Pollard (1987) are incorporated. The extension to any positive integer p comes from extending (1) to cover all p, and then applying the triangle inequality for $\mathcal{L}^p$ norms.

In Section 5, the Chaining Inequality will be applied to a symmetrized version of a degenerate U-process, after first conditioning on certain sources of randomness. The moment inequality presented in Section 4 will provide the means of verifying condition (ii) for this related process. The other three conditions will be easy to verify. This will give us a pth-moment maximal inequality for the conditional, symmetrized process. By averaging out over the conditioning variables we will obtain a corresponding inequality for the unconditional process. We will then translate the latter result into a pth-moment maximal inequality for the degenerate U-process by means of a symmetrization inequality, which is established in the next section.

3. Symmetrization Inequality. Independently, take samples $\{X_i\}_{i=1}^n$ and $\{X_i'\}_{i=1}^n$ from a distribution P on a set $\mathcal{S}$, and a sample $\{\sigma_i\}_{i=1}^n$ from the distribution that assigns probability $\frac{1}{2}$ to each of $+1$ and -1. Call each σ_i a sign variable. Let $\mathcal{F}$ be a class of real-valued functions on $\mathcal{S}^k$, $k \geq 1$. For each $f \in \mathcal{F}$, define

$$(2)\quad \tilde{f}(\boldsymbol{i_k}) = f(X_{i_1}, \ldots, X_{i_k}) \; - \; f(X'_{i_1}, X_{i_2}, \ldots, X_{i_k}) - f(X_{i_1}, X'_{i_2}, \ldots, X_{i_k}) \\ + \; \cdots (-1)^k f(X'_{i_1}, \ldots, X'_{i_k}) \, .$$

That is, $\tilde{f}(\boldsymbol{i_k})$ is a sum of 2^k terms, each having the form

$$(-1)^r f(X^*_{i_1}, \ldots, X^*_{i_k})$$

with the superscipt * equal to a blank (in which case $X^*_i = X_i$) or a prime (in which case $X^*_i = X'_i$), and r is the number of superscripts that are primes.

SYMMETRIZATION INEQUALITY. *Let $Z_1, \ldots, Z_n$ be a sample of independent observations from P, and $\mathcal{F}$ a class of P-degenerate functions on $\mathcal{S}^k$, $k \geq 1$. Let Φ be a convex function on $[0, \infty)$. Then*

$$(3)\qquad \mathbb{P}\Phi\left(\sup_{\mathcal{F}} \left| \sum_{\boldsymbol{i_k}} f(Z_{i_1}, \ldots, Z_{i_k}) \right| \right) \leq \mathbb{P}\Phi\left(\sup_{\mathcal{F}} \left| \sum_{\boldsymbol{i_k}} \sigma_{i_1} \cdots \sigma_{i_k} \tilde{f}(\boldsymbol{i_k}) \right| \right) .$$

PROOF. Consider another sample $Z'_1, \ldots, Z'_n$ from P, where the Z'_i's are independent of the Z_i's. Let $\mathbb{P}_Z$ denote expectation over the Z'_i's, conditional on the Z_i's. Degeneracy implies that $f(Z_{i_1}, \ldots, Z_{i_k})$ can be replaced by a sum of 2^k terms,

$$f(Z_{i_1}, \ldots, Z_{i_k}) \; - \; \mathbb{P}_Z f(Z'_{i_1}, Z_{i_2}, \ldots, Z_{i_k}) - \mathbb{P}_Z f(Z_{i_1}, Z'_{i_2}, \ldots, Z_{i_k}) \\ + \; \cdots (-1)^k \mathbb{P}_Z f(Z'_{i_1}, \ldots, Z'_{i_k})$$

without changing the left-hand side of (3). The pattern is the same as in (2).

The $\mathbb{P}_Z$ can be pulled out of the sum, then past the sup and the Φ by virtue of Jensen's inequality, increasing the quantity on the left-hand side of (3) to

$$(4)\qquad \mathbb{P}\mathbb{P}_Z\Phi\left(\sup_{\mathcal{F}} \left| \sum_{\boldsymbol{i_k}} \left[f(Z_{i_1}, \ldots, Z_{i_k}) - \cdots (-1)^k f(Z'_{i_1}, \ldots, Z'_{i_k}) \right] \right| \right) .$$

Consolidate $\mathbb{P}\mathbb{P}_Z$ into $\mathbb{P}$.

Now suppose the Z_i's and the Z'_i's are constructed in a special way from the double sample $X_1, X'_1, \ldots, X_n, X'_n$ and the σ_i's:

$$(Z_i, Z'_i) = \{\sigma_i = 1\}(X_i, X'_i) + \{\sigma_i = -1\}(X'_i, X_i)$$

That is, σ_i determines the order in which (X_i, X'_i) will be labelled.

Rewrite (4) in terms of the σ_i's, X_i's and X'_i's.

Fix an $\boldsymbol{i_k}$. Concentrate on the contribution from i_1. The sum of 2^k terms in (4) can be grouped as a sum of 2^{k-1} pairs

$$\pm\left[f(Z_{i_1},\ldots)-f(Z'_{i_1},\ldots)\right].$$

Here the terms differ only in the first position. If $\sigma_{i_1}=1$ replace Z_{i_1} by X_{i_1} and Z'_{i_1} by X'_{i_1}; if $\sigma_{i_1}=-1$ replace Z_{i_1} by X'_{i_1} and Z'_{i_1} by X_{i_1}. The term equals

$$\pm\sigma_{i_1}\left[f(X_{i_1},\ldots)-f(X'_{i_1},\ldots)\right].$$

That is, (4) is unchanged if we replace Z_{i_1} by X_{i_1} and Z'_{i_1} by X'_{i_1}, then compensate with a σ_{i_1} factor. In a similar fashion, replace all the other Z_i by X_i and Z'_i by X'_i, transforming (4) into

$$\mathbb{P}\Phi\left(\sup_{\mathcal{F}}\left|\sum_{\boldsymbol{i_k}}\sigma_{i_1}\cdots\sigma_{i_k}\left[f(X_{i_1},\ldots,X_{i_k})-\cdots(-1)^k f(X'_{i_1},\ldots,X'_{i_k})\right]\right|\right).$$

□

Pollard (1982) discovered how to symmetrize using sign variables for the case $k=1$. Nolan and Pollard (1987) developed the technique for $k=2$. The Symmetrization Inequality gives the general result.

For each $f\in\mathcal{F}$, define

$$S_n^k f=(n)_k^{-1}\sum_{\boldsymbol{i_k}}\sigma_{i_1}\cdots\sigma_{i_k}\tilde{f}(\boldsymbol{i_k}).$$

Call the collection $\{S_n^k f : f\in\mathcal{F}\}$ the symmetrized process. The following corollary is obtained by absorbing a scaling factor into f in (3).

Corollary 2. *Let $\mathcal{F}$ be a class of P-degenerate functions on $\mathcal{S}^k$, $k\geq 1$. Let Φ be a convex function on $[0,\infty)$. Then*

$$\mathbb{P}\Phi\left(\sup_{\mathcal{F}}|n^{k/2}U_n^k f|\right)\leq\mathbb{P}\Phi\left(\sup_{\mathcal{F}}|n^{k/2}S_n^k f|\right).$$

□

Notice that $S_n^k f$, conditioned on the $\{X_i\}$ and the $\{X'_i\}$, is just a kth-order form in the sign variables. In the next section, we present a moment inequality for kth-order forms having this simple structure. This inequality will enable us to establish condition (ii) of the Chaining Inequality for the symmetrized process, conditioned on the two samples.

4. Moment Inequality. For $k\geq 1$, let $a(i_1,\ldots,i_k)$ be a real-valued function of k indices, each running from 1 to n, and let $\sigma_1,\ldots,\sigma_n$ be a sequence of independent sign variables. Suppose $a(i_1,\ldots,i_k)=0$ whenever two indices are equal, and let $\|a\|_2$ denote the ℓ^2 norm $(\sum_{\boldsymbol{i_k}} a(i_1,\ldots,i_k)^2)^{1/2}$.

Moment Inequality. *For each positive integer q there is a constant $C(k,q)$ such that*

$$\mathbb{P}\Bigg(\sum_{\boldsymbol{i_k}} \sigma_{i_1}\cdots\sigma_{i_k} a(i_1,\ldots,i_k)\Bigg)^q \le C(k,q)\|a\|_2^q\,.$$

Bonami (1970, Chapitre II, Théorème 6) appears to have been the first to establish this type of result. She also obtained an explicit bound for the constant $C(k,q)$. For the purposes of this paper, what is crucial is not the form of $C(k,q)$, but rather the fact that it does not depend on n. See Sherman (1991b) for a simple proof of this inequality.

5. Maximal Inequality. Recall the definition of $\tilde f(\boldsymbol{i_k})$ given in (2). For $f,g\in\mathcal{F}$, let $h=f-g$ and define $\tilde h(\boldsymbol{i_k})=\tilde f(\boldsymbol{i_k})-\tilde g(\boldsymbol{i_k})$. For $i=1,\ldots,n$, define $W_i=X_i$ and $W_{n+i}=X_i'$. Use the fact that $\tilde h(\boldsymbol{i_k})$ is a sum of 2^k terms, and make k applications of the inequality $(a+b)^2\le 2(a^2+b^2)$, for real numbers a and b, to see that

$$\sum_{\boldsymbol{i_k}} \tilde h(\boldsymbol{i_k})^2 \le 2^k \sum_{\boldsymbol{j_k}} h(W_{j_1},\ldots,W_{j_k})^2 \tag{5}$$

where $\boldsymbol{j_k}=(j_1,\ldots,j_k)$ ranges over the $(2n)_k$ k-tuples of distinct integers from the set $\{1,\ldots,2n\}$.

Let F be an envelope for $\mathcal{F}$. That is, $\sup_{\mathcal{F}}|f(\cdot)|\le F(\cdot)$ for all f in $\mathcal{F}$. Write U_{2n}^k for the probability measure putting mass $(2n)_k^{-1}$ on each of the $(2n)_k$ k-tuples $(W_{j_1},\ldots,W_{j_k})$. Assume $U_{2n}^kF^2<\infty$. For $f,g\in\mathcal{F}$, define the pseudometric

$$d_{U_{2n}^k}(f,g)=\Big[U_{2n}^k|f-g|^2/U_{2n}^kF^2\Big]^{1/2}.$$

Note that $d_{U_{2n}^k}$ depends on the envelope F. Since this dependence is not crucial to what follows, it is suppressed in the notation.

We are now prepared to prove the main result of this paper. We do so by applying the Chaining Inequality to the symmetrized process $\{S_n^kf : f\in\mathcal{F}\}$, conditioning at first on the double sample $W_1,\ldots,W_{2n}$.

Maximal Inequality. *Let $\mathcal{F}$ be a class of P-degenerate functions on $\mathcal{S}^k$, $k\ge 1$. Let F be an envelope for $\mathcal{F}$, and let p and m be positive integers. If $P^kF^2<\infty$, where P^k is the product measure $P\otimes\cdots\otimes P$ (k factors), then*

$$\mathbb{P}\sup_{\mathcal{F}}\left|n^{k/2}U_n^kf\right|^p \le \Gamma\mathbb{P}\left[\tau_n^k\int_0^{\delta_n^k}\Big[D(x,d_{U_{2n}^k},\mathcal{F})\Big]^{1/2mp}\,dx\right]^p$$

where Γ *is a universal constant,* $\tau_n^k = \sqrt{U_{2n}^k F^2}$*, and* $\delta_n^k \tau_n^k = \sup_{\mathcal{F}} \sqrt{U_{2n}^k f^2}$.

PROOF. Let γ denote a constant to be determined shortly. Define

$$\begin{aligned} Z(f) &= n^{k/2} S_n^k f/\tau_n^k\,, \\ \Psi(x) &= x^{2m}/\gamma\,, \\ d(f,g) &= d_{U_{2n}^k}(f,g)\,. \end{aligned}$$

Write $\mathbb{P}_W$ for expectation conditional on $W = (W_1, \ldots, W_{2n})$. Check condition (ii) of the Chaining Inequality. Remember that $h = f - g$.

$$\begin{aligned} \mathbb{P}_W \Psi_p(|Z(f) - Z(g)|/d(f,g)) &= \gamma^{-1} \mathbb{P}_W \left(n^{k/2} S_n^k h / \left(U_{2n}^k h^2\right)^{1/2} \right)^{2mp} \\ &= \gamma^{-1} \mathbb{P}_W \left(\sum_{\boldsymbol{i_k}} \sigma_{i_1} \ldots \sigma_{i_k} a(i_1, \ldots, i_k) \right)^{2mp} \end{aligned}$$

where

$$a(i_1, \ldots, i_k) = \left[n^{k/2} (2n)_k^{1/2} (n)_k^{-1}\, \tilde{h}(\boldsymbol{i_k}) \right] / \left(\sum_{\boldsymbol{j_k}} h(W_{j_1}, \ldots, W_{j_k})^2 \right)^{1/2}.$$

Apply the Moment Inequality to bound the last conditional expectation by $\gamma^{-1} C(k, 2mp) \|a\|_2^{2mp}$. Eventually, $n^{k/2}(2n)_k^{1/2}(n)_k^{-1}$ is bounded by 2^k. Deduce from this and (5) that $\|a\|_2 \le 2^{k+k/2}$. Set γ equal to $C(k, 2mp)(2^{k+k/2})^{2mp}$ to establish condition (ii) of the Chaining Inequality. Take t_0 to be the zero function to verify condition (iii). The other two conditions are trivially satisfied. Take pth powers to see that

$$\mathbb{P}_W \sup_{\mathcal{F}} \left| n^{k/2} S_n^k f / \tau_n^k \right|^p \le \left[2 \int_0^{\delta_n^k} \left[\gamma D(x, d_{U_{2n}^k}, \mathcal{F}) \right]^{1/2mp} dx \right]^p.$$

Multiply through by $|\tau_n^k|^p$, take expectations, and let $\Gamma = 2^p \gamma^{1/2m}$ to get

$$\mathbb{P} \sup_{\mathcal{F}} \left| n^{k/2} S_n^k f \right|^p \le \Gamma \mathbb{P} \left[\tau_n^k \int_0^{\delta_n^k} \left[D(x, d_{U_{2n}^k}, \mathcal{F}) \right]^{1/2mp} dx \right]^p.$$

Appeal to Corollary 2 with $\Phi(x) = |x|^p$ to complete the proof. □

6. Consequences. For the Maximal Inequality to be useful in practice, there must exist a function that dominates $D(x, d_{U_{2n}^k}, \mathcal{F})$ on $(0, 1]$, and whose $2mp$th root is integrable on $(0, 1]$. Moreover, this dominating function should not depend on n. When $\mathcal{F}$ satisfies a mild regularity condition called a Euclidean condition, these requirements are satisfied.

DEFINITION 3. Let $\mathcal{F}$ be a class of real-valued functions on a set $\mathcal{X}$. Call $\mathcal{F}$ *Euclidean* for the envelope F if there exist positive constants A and V with the following property: if μ is a measure for which $\mu F^2 < \infty$, then

$$D(x, d_\mu, \mathcal{F}) \le Ax^{-V} \quad 0 < x \le 1 \,,$$

where, for $f, g \in \mathcal{F}$,

$$d_\mu(f,g) = \left[\mu|f-g|^2/\mu F^2\right]^{1/2} .$$

The constants A and V must not depend on μ. We shall also say that $\mathcal{F}$ is Euclidean(A, V) for the envelope F.

Implicit in the definition is the assumption that the functions comprising $\mathcal{F}$ and the envelope F are μ-measurable with respect to a fixed σ-field on $\mathcal{X}$. Simple criteria exist for determining the Euclidean property. Nolan and Pollard (1987) collect a number of such criteria. Pakes and Pollard (1989) provide complementary results.

MAIN COROLLARY. *Let $\mathcal{F}$ be a class of P-degenerate functions on $\mathcal{S}^k$, $k \ge 1$, and let P^k denote the product measure $P \otimes \cdots \otimes P$ (k factors). Suppose $\mathcal{F}$ is Euclidean(A,V) for an envelope F satisfying $P^kF^2 < \infty$. For a given positive integer p and $0 < \epsilon < 1$, choose a positive integer m large enough so that $(1 - V/2mp)p > p - \epsilon$. Then*

$$I\!\!P \sup_{\mathcal{F}} \left|n^{k/2} U_n^k f\right|^p \le \Lambda \left[I\!\!P \sup_{\mathcal{F}} (U_{2n}^k f^2)^\alpha \right]^{1/2}$$

where Λ is a universal constant and $\alpha = (1 - V/2mp)p$.

PROOF. Substitute Ax^{-V} for $D(x, d_{U_{2n}^k}, \mathcal{F})$ in the upper bound of the Maximal Inequality. Integrate, take expectations, and then apply the Cauchy-Schwarz inequality. The resulting bound equals

$$\Gamma A^{1/2mp} \left[I\!\!P(U_{2n}^k F^2)^{p-\alpha} \right]^{1/2} \left[I\!\!P \sup_{\mathcal{F}} (U_{2n}^k f^2)^\alpha \right]^{1/2} .$$

Since $0 < p - \alpha < \epsilon < 1$,

$$I\!\!P(U_{2n}^k F^2)^{p-\alpha} \le (P^k F^2)^\epsilon < \infty \,.$$

Take $\Lambda = \Gamma A^{1/2mp} (P^k F^2)^{\epsilon/2}$ to complete the proof. □

The rest of this section is devoted to drawing useful consequences from the Main Corollary.

COROLLARY 4. *Suppose the conditions of the Main Corollary hold. Then*

(i) $I\!\!P \sup_{\mathcal{F}} |n^{k/2} U_n^k f| = O(1)$

(ii) $\sup_{\mathcal{F}} |n^{k/2} U_n^k f| = O_p(1)$

PROOF. Apply the Main Corollary with $p = 1$. Then $0 < \alpha < 1$, and so

$$I\!\!P \sup_{\mathcal{F}} (U_{2n}^k f^2)^\alpha \leq I\!\!P (U_{2n}^k F^2)^\alpha \leq (P^k F^2)^\alpha < \infty \,.$$

This establishes (i). Chebyshev's inequality turns (i) into (ii). □

COROLLARY 4A. *Suppose the conditions of the Main Corollary hold, and that $p \geq 2$. If, in addition, $P^k F^{4p} < \infty$, then*

$$I\!\!P \sup_{\mathcal{F}} |n^{k/2} U_n^k f|^p = O(1) \,.$$

PROOF. Apply the Main Corollary again. By the Cauchy-Schwarz inequality,

$$I\!\!P \sup_{\mathcal{F}} (U_{2n}^k f^2)^\alpha \leq \left[I\!\!P (U_{2n}^k F^2)^{2\alpha} \right]^{1/2} \,.$$

Since $p \geq 2$, $\alpha \geq 1$ and so $2\alpha \geq 2$. Also, $P^k F^{4p} < \infty$ implies $P^k F^{4\alpha} < \infty$. Therefore, we may apply Lemma A of Serfling (1980, p.185) to get that $U_{2n}^k F^2$ converges to $P^k F^2$ in rth mean, where $r = 2\alpha$. It then follows from Theorem B of Serfling (1980, p.15) that as n tends to infinity,

$$I\!\!P (U_{2n}^k F^2)^{2\alpha} \longrightarrow (P^k F^2)^{2\alpha} < \infty \,,$$

from which the result follows. □

We now present a simple decomposition of a U-statistic (process) of order k into a sum of degenerate U-statistics (processes). This decomposition will come in handy when we analyze the rank estimator in Section 7.

Recall that P denotes the sampling distribution on a set $\mathcal{S}$. Let $\mathcal{F}$ be a class of real-valued functions on $\mathcal{S}^k$, $k \geq 1$. Fix $f \in \mathcal{F}$. Suppose $P^k f < \infty$, where P^k is the product measure $P \otimes \cdots \otimes P$ (k factors). Then there exist functions $f_1, \ldots, f_k$ such that for each i, f_i is P-degenerate on $\mathcal{S}^i$ and

$$U_n^k f = P^k f + P_n f_1 + \sum_{i=2}^{k} U_n^i f_i \,. \tag{6}$$

Moreover, for each $s \in \mathcal{S}$,

$$f_1(s) = f(s, P, \ldots, P) + \cdots + f(P, \ldots, P, s) - k P^k f \,. \tag{7}$$

The notation, $f(s, P \ldots, P)$, for example, is short for the conditional expectation, under P^k, of f given its first argument. Special attention is given to f_1

because $P_n f_1$ is the dominant stochastic term in the decomposition when f_1 is not identically zero. This fact is crucial to the application in Section 8. The proof of (6) and (7) is straightforward. Serfling (1980, pp.177-178) gives details.

The next result is a straightforward extension of Lemma 20 in Nolan and Pollard (1987).

LEMMA 5. *Let $\mathcal{S}$ be a set and $\mathcal{F}$ a class of real-valued functions on $\mathcal{S}^k$, for $k \geq 1$. Let ν be a probability measure on $\mathcal{S}$. If $\mathcal{F}$ is Euclidean for an envelope F, then the class $\{\nu f(s_1, \ldots, s_{i-1}, \cdot, s_{i+1}, \ldots, s_k) : f \in \mathcal{F}\}$ is Euclidean for the envelope $\sqrt{\nu F^2}$, $i = 1, \ldots, k$.* □

Refer to (6), and let $\mathcal{F}_i = \{f_i : f \in \mathcal{F}\}$.

LEMMA 6. *If $\mathcal{F}$ is Euclidean for an envelope F satisfying $P^k F^2 < \infty$, then $\mathcal{F}_i$ is Euclidean for an envelope F_i satisfying $P^i F_i^2 < \infty$, $i = 1, \ldots, k$.*

PROOF. Fix f_i in $\mathcal{F}_i$ and consider the corresponding f in $\mathcal{F}$. The proof of (6) given by Serfling (1980, pp.177-178) shows that f_i is a sum of a finite number of terms, each of which is plus or minus a conditional expectation of f given j of its arguments, where $j \in \{0, \ldots, i\}$. The result follows from repeated application of Lemma 5 above and Lemma 16 in Nolan and Pollard (1987). □

COROLLARY 7. *Let $\mathcal{F}$ be a class of zero-mean functions on $\mathcal{S}^k$, $k \geq 1$. If $\mathcal{F}$ is Euclidean for an envelope F satisfying $P^k F^2 < \infty$, then*

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

PROOF. Recall that $P_n \equiv U_n^1$. Deduce from (6) that

$$\sup_{\mathcal{F}} |U_n^k f| \leq \sum_{i=1}^{k} \sup_{\mathcal{F}_i} |U_n^i f_i| \,.$$

By Lemma 6, $\mathcal{F}_i$ is Euclidean for an envelope F_i satisfying $P^i F_i^2 < \infty$. By Corollary 4(ii),

$$\sup_{\mathcal{F}_i} |U_n^i f_i| = O_p(n^{-i/2}) = O_p(1/\sqrt{n}) \,.$$

□

The next result applies to a class of functions of the form $\{f(\cdot, \theta) : \theta \in \Theta\}$ where Θ is a subset of $\mathbb{R}^d$. It provides a more delicate uniformity result useful for establishing the asymptotic normality of an estimator defined by optimization of a random criterion function of U-process structure. To establish this result, we must dig a bit deeper into the bound given in the Main Corollary.

Corollary 8. *Let $\mathcal{F}$ be a class of P-degenerate functions on $\mathcal{S}^k$, $k \geq 1$. Suppose $\mathcal{F}$ has the form $\{f(\cdot,\theta) : \theta \in \Theta\}$, and that θ_0 is a point in Θ for which $f(\cdot,\theta_0) \equiv 0$. Let $P^k = P \otimes \cdots \otimes P$ (k factors). If*

(i) *$\mathcal{F}$ is Euclidean for an envelope F satisfying $P^k F^2 < \infty$,*

(ii) $P^k|f(\cdot,\theta)| \to 0 \quad \text{as} \quad \theta \to \theta_0$,

then uniformly over $o_p(1)$ neighborhoods of θ_0,

$$U_n^k f(\cdot,\theta) = o_p(1/n^{k/2})\,.$$

Proof. Let $\{\epsilon_n\}$ be a sequence of nonnegative real numbers converging to zero and $\Theta_n = \{\theta \in \Theta : |\theta - \theta_0| \leq \epsilon_n\}$. The result is equivalent to

$$\sup_{\Theta_n} |n^{k/2} U_n^k f(\cdot,\theta)| = o_p(1)\,. \tag{8}$$

Apply the Main Corollary with $p = 1$ to get

$$\mathbb{P} \sup_{\Theta_n} |n^{k/2} U_n^k f(\cdot,\theta)| \leq \Lambda \left[\mathbb{P} \sup_{\Theta_n} (U_{2n}^k f(\cdot,\theta)^2)^\alpha\right]^{1/2}.$$

Since $0 < \alpha < 1$,

$$\mathbb{P} \sup_{\Theta_n} \left(U_{2n}^k f(\cdot,\theta)^2\right)^\alpha \leq \left[\mathbb{P} \sup_{\Theta_n} U_{2n}^k f(\cdot,\theta)^2\right]^\alpha.$$

We now show that the bound in the last inequality has order $o(1)$. This, combined with Chebyshev's inequality, will establish (8).

Given $\epsilon > 0$, choose a constant M so large that $P^k F^2\{F > M\} < \epsilon$. Then

$$\mathbb{P} \sup_{\Theta_n} U_{2n}^k f(\cdot,\theta)^2 \leq \epsilon + M \mathbb{P} \sup_{\mathcal{F}} U_{2n}^k |f(\cdot,\theta)|\,.$$

Note that $\{U_{2n}^k |f(\cdot,\theta)| : \theta \in \Theta_n\}$ is a U-process of order k. Deduce from (6) that

$$\mathbb{P} \sup_{\Theta_n} U_{2n}^k |f(\cdot,\theta)| \leq \sup_{\Theta_n} P^k |f(\cdot,\theta)| + \sum_{i=1}^k \mathbb{P} \sup_{\Theta_n} |U_{2n}^i g_i(\cdot,\theta)|$$

where the class of functions $\{g_i(\cdot,\theta) : \theta \in \Theta\}$ is P-degenerate on $\mathcal{S}^i$.

Deduce from Lemma 2.14(iii) in Pakes and Pollard (1989) that the class $\{|f(\cdot,\theta)| : \theta \in \Theta\}$ is Euclidean for the envelope F. By Lemma 6, the class $\{g_i(\cdot,\theta) : \theta \in \Theta\}$ is Euclidean for an envelope F_i satisfying $P^i F_i^2 < \infty$. Apply Corollary 4(i) to see that

$$\mathbb{P} \sup_{\Theta} |U_{2n}^i g_i(\cdot,\theta)| = O(n^{-i/2}) = o(1)\,.$$

By assumption (ii), $\sup_{\Theta_n} P^k|f(\cdot,\theta)| = o(1)$. This proves (8). □

The final result provides rates of uniform almost-sure convergence.

COROLLARY 9. *Suppose all the conditions of the Main Corollary hold. For real numbers $\delta > 0$ and $\beta > 1$, let p be a positive integer satisfying $p \geq \beta/\delta$. If $P^k F^{4p} < \infty$, then*

$$\sup_{\mathcal{F}} \left| n^{k/2-\delta} U_n^k f \right| \to 0$$

almost surely as n tends to infinity.

PROOF. By Corollary 4A, $\mathbb{P} \sup_{\mathcal{F}} |n^{k/2} U_n^k f|^p \leq M < \infty$. For each $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}\left\{ \sup_{\mathcal{F}} \left| n^{k/2-\delta} U_n^k f \right| > \epsilon \right\} &\leq \mathbb{P} \sup_{\mathcal{F}} \left| n^{k/2} U_n^k f \right|^p / \epsilon^p n^{\delta p} \\ &\leq Cn^{-\beta} \end{aligned}$$

where $C = M/\epsilon^p$. Apply the Borel-Cantelli lemma to complete the proof. □

7. A Semiparametric Rank Estimator. Let $Z = (Y, X)$ be an observation from a distribution P on a set $\mathcal{S} \subseteq \mathbb{R} \otimes \mathbb{R}^d$, where Y is a response variable, and X is a vector of regressors. Han (1987) introduced the generalized regression model

$$Y = D \circ F(X'\beta_0, u)$$

where β_0 is a d-dimensional vector of unknown parameters, u is a random variable independent of X, F is a strictly increasing function of each of its arguments, and D is a monotone increasing function of its argument. Many interesting regression models fit into this framework. For example, take $F(x,y) = x+y$. If $D(z) = z$, the model reduces to a standard linear regression model; for $D(z) = \{z \geq 0\}$, a binary choice model; for $D(z) = z\{z \geq 0\}$, a censored regression model. Transformation and duration models are other special cases.

Let $Z_1, \ldots, Z_n$ be a sample of independent observations from P. Cavanagh (1990) proposed estimating β_0 with $\beta_n = \operatorname{argmax}_{\boldsymbol{R}^d} G_n(\beta)$ where

$$G_n(\beta) = (n)_3^{-1} \sum_{\boldsymbol{i_3}} \{Y_i > Y_j\}\{X_i'\beta > X_k'\beta\}\,. \tag{9}$$

Here $\boldsymbol{i_3} = (i,j,k)$ ranges over the $(n)_3$ ordered triples of distinct integers from the set $\{1,\ldots,n\}$. Note that $\{G_n(\beta) : \beta \in \mathbb{R}^d\}$ is a U-process of order three.

In order to motivate the estimator, let $\{a_1,\ldots,a_n\}$ be a set of real numbers and let $R_n(a_i)$ denote the rank of a_i. The monotonicity of $D \circ F$ and the independence of the u_i's and X_i's ensure that

$$R_n(\mathbb{P}(Y_i \mid X_i)) = R_n(X_i'\beta_0) \tag{10}$$

where possible ties are handled in an obvious way. Let $\sigma(1), \ldots, \sigma(n)$ denote a permutation of the set $\{1, \ldots, n\}$, and note that $\sum_i i\,\sigma(i)$ is maximized when $\sigma(i) = i$. This and (10) suggest estimating β_0 with the maximizer of

$$\sum_i R_n(Y_i) R_n(X_i'\beta)\,.$$

The facts $R_n(Y_i) = \sum_j \{Y_i \geq Y_j\}$ and $R_n(X_i'\beta) = \sum_k \{X_i'\beta \geq X_k'\beta\}$ lead to the proposed estimator β_n. Terms involving ties or equal indices turn out not to matter and so are discarded from the criterion function in (9).

Notice that $G_n(\beta)$ is a discontinuous function of β. Standard methods for determining the asymptotic distribution of an optimization estimator require some form of smoothness on the criterion function and so do not apply. In the next subsection, we present a method that is general enough to cover Cavanagh's estimator.

7.1 A General Method. Let Θ be a subset of $\mathbb{R}^m$, and θ_0 an element of Θ and a parameter of interest. Suppose θ_0 maximizes a function $\Gamma(\theta)$ defined on Θ. Suppose further that a sample analogue, $\Gamma_n(\theta)$, is maximized at a point θ_n that converges in probability to θ_0.

In this section, we present a general method for establishing that θ_n is $\sqrt{n}$-consistent for θ_0 and asymptotically normally distributed. This method has its origins in a paper by Huber (1967) and has been recast into the form presented here (apart from minor modifications) by Pollard (1989a). The method is embodied in the two theorems that follow. The first of these provides conditions under which θ_n is $\sqrt{n}$-consistent for θ_0. The second theorem gives conditions under which a $\sqrt{n}$-consistent estimator is also asymptotically normally distributed. The proofs appear in Sherman (1992).

For simplicity, we will assume that θ_0 is the zero vector (denoted $\mathbf{0}$) in $\mathbb{R}^m$, and that $\Gamma_n(\theta_0) = \Gamma(\theta_0) = 0$. This can always be arranged by working with $\Gamma_n(\theta_0 + t) - \Gamma_n(\theta_0)$ instead of $\Gamma_n(\theta)$, $\Gamma(\theta_0 + t) - \Gamma(\theta_0)$ instead of $\Gamma(\theta)$, and substituting t for θ, where t satisfies $\theta_0 + t \in \Theta$.

THEOREM 1: *Let θ_n be a maximizer of $\Gamma_n(\theta)$, and $\mathbf{0}$ a maximizer of $\Gamma(\theta)$. Suppose θ_n converges in probability to $\mathbf{0}$, and also that*

(i) *there exists a neighborhood $\mathcal{N}$ of $\mathbf{0}$ and a constant $\kappa > 0$ for which*

$$\Gamma(\theta) \leq -\kappa|\theta|^2$$

for all θ in $\mathcal{N}$;

(ii) *uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$,*

$$\Gamma_n(\theta) = \Gamma(\theta) + O_p(|\theta|/\sqrt{n}) + o_p(|\theta|^2) + O_p(1/n)\,.$$

Then

$$|\theta_n| = O_p(1/\sqrt{n})\,.$$

Once $\sqrt{n}$-consistency of θ_n is established, we can prove asymptotic normality provided there exist very good quadratic approximations to $\Gamma_n(\theta)$ within $O_p(1/\sqrt{n})$ neighborhoods of $\mathbf{0}$. In the following theorem, the symbol $\Longrightarrow$ denotes convergence in distribution.

THEOREM 2: *Suppose θ_n is $\sqrt{n}$-consistent for $\mathbf{0}$, an interior point of Θ. Suppose also that uniformly over $O_p(1/\sqrt{n})$ neighborhoods of $\mathbf{0}$,*

$$\Gamma_n(\theta) = \frac{1}{2}\theta' V\theta + \frac{1}{\sqrt{n}}\theta' W_n + o_p(1/n) \tag{11}$$

where V is a negative definite matrix, and W_n converges in distribution to a $N(\mathbf{0}, \Delta)$ random vector. Then

$$\sqrt{n}\theta_n \Longrightarrow N(\mathbf{0}, V^{-1}\Delta V^{-1})\,.$$

The conditions of the theorems do not require that θ_n be a zero of the gradient of $\Gamma_n(\theta)$. Nor do they require that $\Gamma_n(\theta)$ be a continuous function of θ. This approach, then, provides a framework within which the asymptotic distribution of Cavanagh's estimator can be established.

7.2. Consistency. Notice that if β_n maximizes $G_n(\beta)$, then so does $c\beta_n$, for any $c > 0$, since $G_n(\beta) = G_n(c\beta)$ for any $c > 0$. In order to achieve a unique parametrization, restrict the parameter space to a subset of $\{\beta \in \mathbb{R}^d : \beta_d = 1\}$. That is, assume that one element of β_0 is known to be nonzero, and normalize the parameter space by this value. Let $\mathcal{B}$ denote this restricted parameter space. Rather than introduce new notation, rechristen β_n as the maximizer of $G_n(\beta)$ over $\mathcal{B}$.

We now state the assumptions used to prove the consistency of β_n.

A1. The distribution of the regressor X has a continuous density with respect to Lebesgue measure on $\mathbb{R}^d$.
A2. The function $H(t) = \mathbb{P}(\{Y_1 > Y_2\} \mid X_1'\beta_0 = t)$ is strictly increasing.
A3. The parameter space $\mathcal{B}$ is compact.
A4. The function $G(\beta) = \mathbb{P}\{Y_1 > Y_2\}\{X_1'\beta > X_3'\beta\}$ is continuous on $\mathcal{B}$.

These assumptions are much stronger than necessary, but are made to simplify the exposition. Notice that $G(\beta)$ is the expected value of $G_n(\beta)$.

CONSISTENCY. *If A1 through A4 hold, then* $|\beta_n - \beta_0| = o_p(1)$.

PROOF. We will show

(i) $G(\beta)$ is uniquely maximized at β_0.

(ii) $\sup_{\mathcal{B}} |G_n(\beta) - G(\beta)| = o_p(1)$.

Consistency then follows from standard arguments using A3 and A4. (See, for example, Amemiya (1985, pp.106–107).)

By symmetry,

$$G(\beta) = \frac{1}{2} I\!P[H(X_1'\beta_0)\{X_1'\beta > X_3'\beta\} + H(X_3'\beta_0)\{X_1'\beta < X_3'\beta\}]\,. \tag{12}$$

If $\beta = \beta_0$, then A1 and A2 ensure that the indicators in (12) pick out the larger of $H(X_1'\beta_0)$ and $H(X_3'\beta_0)$ with probability one. Consequently,

$$G(\beta_0) = \frac{1}{2} I\!P \max\left(H(X_1'\beta_0), H(X_3'\beta_0)\right)\,.$$

Deduce that $G(\beta)$ is maximized at β_0.

Suppose that for some $\beta \neq \beta_0$,

$$G(\beta) = \frac{1}{2} I\!P \max\left(H(X_1'\beta_0), H(X_3'\beta_0)\right)\,. \tag{13}$$

Deduce from (13) and (12) that

$$H(X_1'\beta_0) \geq H(X_3'\beta_0) \quad \text{when} \quad X_1'\beta > X_3'\beta\,. \tag{14}$$

Let $W = X_1 - X_3$. Deduce from (14) and A2 that

$$W'\beta_0 \geq 0 \quad \text{when} \quad W'\beta > 0\,. \tag{15}$$

Finally, from (15), deduce that

$$I\!P\{W \in D\} = 0 \tag{16}$$

where

$$D = \{W'\beta_0 < 0\}\{W'\beta > 0\}\,.$$

From A1, the distribution of W has a continuous density with respect to Lebesgue measure on $I\!R^d$ given by

$$\rho(w) = \int_{\boldsymbol{R}^d} f(x) f(w+x)\, dx$$

where f is the density function of the distribution of X. Note that $\rho(\mathbf{0}) > 0$. From this and the continuity of ρ deduce that ρ is bounded away from zero in a neighborhood of the origin. Since $\beta \neq \beta_0$, D is a d-dimensional subset of $\mathbb{R}^d$ intersecting every neighborhood of the origin. But then $\mathbb{P}\{W \in D\} > 0$, contradicting (16). This proves (i).

We prove a much stronger result than (ii) without extra effort. For each (z_1, z_2, z_3) in $\mathcal{S}^3 = \mathcal{S} \otimes \mathcal{S} \otimes \mathcal{S}$ and each β in $\mathcal{B}$, let

$$f(z_1, z_2, z_3, \beta) = \{y_1 > y_2\}\{x_1'\beta > x_3{}'\beta\} - G(\beta)\,.$$

Then

$$G_n(\beta) - G(\beta) = U_n^3 f(\cdot, \beta)\,.$$

The class of functions $\{f(\cdot, \beta) : \beta \in \mathcal{B}\}$ is shown in Subsection 7.4 to be Euclidean for the constant envelope 1. Since $\{U_n^3 f(\cdot, \beta) : \beta \in \mathcal{B}\}$ is a zero-mean U-process of order three, it follows from Corollary 7 that

$$\sup_{\mathcal{B}} |U_n^3 f(\cdot, \beta)| = O_p(1/\sqrt{n})\,.$$

This is more than enough to prove (ii). □

7.3. $\sqrt{n}$-Consistency and Asymptotic Normality. Represent each β in $\mathcal{B}$ as $\beta(\theta) = (\theta, 1)$ where θ is an element of Θ, a compact subset of $\mathbb{R}^{d-1}$. Also, write $\beta_0 = \beta(\theta_0)$ where θ_0 consists of the first $d-1$ components of β_0. Let θ_n denotes the first $d-1$ components of β_n. The consistency of β_n for β_0 immediately implies the consistency of θ_n for θ_0.

For each θ in Θ, write $\Gamma(\theta)$ for $G(\beta(\theta)) - G(\beta(\theta_0))$. Similarly, write $\Gamma_n(\theta)$ for $G_n(\beta(\theta)) - G_n(\beta(\theta_0))$. Since $G(\beta)$ is maximized at β_0, $\Gamma(\theta)$ is maximized at θ_0. Similarly, θ_n maximizes $\Gamma_n(\theta)$ over Θ. As in Section 2, we shall assume that $\theta_0 = \mathbf{0}$, the zero vector in $\mathbb{R}^{d-1}$. Thus, $\Gamma_n(\mathbf{0}) = \Gamma(\mathbf{0}) = 0$.

For each (z_1, z_2, z_3) in $\mathcal{S}^3$ and each θ in Θ, define

$$h(z_1, z_2, z_3, \theta) = \{y_1 > y_2\}\{x_1'\beta(\theta) > x_3'\beta(\theta)\}\,. \tag{17}$$

For each z in $\mathcal{S}$ and each θ in Θ, define

$$\tau(z, \theta) = h(z, P, P, \theta) + h(P, z, P, \theta) + h(P, P, z, \theta) \tag{18}$$

Recall from (7) that $h(z, P, P, \theta)$, for example, denotes the conditional expectation of $h(\cdot, \theta)$ given its first argument. The function $\tau(Z_i, \theta)$ will be the ith summand of the empirical process that drives the asymptotic behavior of θ_n. Notice that even though $h(z_1, z_2, z_3, \cdot)$ is discontinuous, $\tau(z, \cdot)$ can be many times differentiable provided the distribution of $X'\beta(\theta)$ is sufficiently smooth.

Write ∇_m for the mth partial derivative operator with respect to θ, and

$$|\nabla_m|\sigma(\theta) \equiv \sum_{i_1, \ldots, i_m} \left| \frac{\partial^m}{\partial\theta_{i_1} \cdots \partial\theta_{i_m}} \sigma(\theta) \right|\,.$$

The symbol $\|\cdot\|$ denotes the matrix norm: $\|(a_{ij})\| = (\sum_{i,j} a_{ij}^2)^{1/2}$.

We now state the last assumption used in the normality proof for θ_n.

A5. Let $\mathcal{N}$ denote a neighborhood of $\mathbf{0}$.

(i) For each z in S, all mixed second partial derivatives of $\tau(z,\cdot)$ exist on $\mathcal{N}$.

(ii) There is an integrable function $M(z)$ such that for all z in S and θ in $\mathcal{N}$

$$\|\nabla_2\tau(z,\theta) - \nabla_2\tau(z,\mathbf{0})\| \le M(z)|\theta| .$$

(iii) $P|\nabla_1\tau(\cdot,\mathbf{0})|^2 < \infty$.

(iv) $P|\nabla_2|\tau(\cdot,\mathbf{0}) < \infty$.

(v) The matrix $P\nabla_2\tau(\cdot,\mathbf{0})$ is negative definite.

The conditions of A5 are standard regularity conditions sufficient to support an argument based on a Taylor expansion of $\tau(\cdot,\theta)$ about $\mathbf{0}$.

Asymptotic Normality. *If A1 through A4 hold, then*

$$\sqrt{n}\theta_n \Longrightarrow N(\mathbf{0}, V^{-1}\Delta V^{-1})$$

where $3V = P\nabla_2\tau(\cdot,\mathbf{0})$ *and* $\Delta = P\nabla_1\tau(\cdot,\mathbf{0})[\nabla_1\tau(\cdot,\mathbf{0})]'$.

Proof. We will show that

$$\Gamma_n(\theta) = \frac{1}{2}\theta' V\theta + \frac{1}{\sqrt{n}}\theta' W_n + o_p(|\theta|^2) + o_p(1/n) \tag{19}$$

uniformly in $o_p(1)$ neighborhoods of $\mathbf{0}$, where W_n converges in distribution to a $N(\mathbf{0},\Delta)$ random vector. Since V is, by A5(v), a negative definite matrix, it will follow from (19) and Theorem 17 that

$$|\theta_n| = O_p(1/\sqrt{n}) . \tag{20}$$

The result can then be deduced from equations (19) and (20), and Theorem 18.

Recall the definition of $h(z_1,z_2,z_3,\theta)$ given in (17). Define

$$f(z_1,z_2,z_3,\theta) = h(z_1,z_2,z_3,\theta) - h(z_1,z_2,z_3,\mathbf{0}) .$$

Since $\Gamma_n(\theta)$ is a U-statistic of order three with expectation $\Gamma(\theta)$, we may apply the decomposition in (6) to write

$$\Gamma_n(\theta) = \Gamma(\theta) + P_n f_1(\cdot,\theta) + U_n^2 f_2(\cdot,\theta) + U_n^3 f_3(\cdot,\theta) \tag{21}$$

where

$$f_1(z,\theta) = f(z,P,P,\theta) + f(P,z,P,\theta) + f(P,P,z,\theta)$$

and, for each θ in Θ, $f_i(\cdot,\theta)$ is P-degenerate on $\mathcal{S}^i$, $i = 2, 3$.

Standard arguments (see, for example, Sherman (1993)) based on A5, a two-term Taylor expansion of $\tau(\cdot,\theta)$ about $\mathbf{0}$, and the fact that $P\tau(\cdot,\theta) = 3\Gamma(\theta)$ show that

$$\Gamma(\theta) = \frac{1}{2}\theta' V\theta + o(|\theta|^2) \quad \text{as} \quad \theta \to \mathbf{0} \tag{22}$$

and

$$P_n f_1(\cdot,\theta) = \frac{1}{\sqrt{n}}\theta' W_n + o_p(|\theta|^2) \tag{23}$$

uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$, where $W_n = \sqrt{n}P_n\nabla_1\tau(\cdot,\mathbf{0})$.

In order to establish (19), it remains to show that

$$U_n^2 f_2(\cdot,\theta) + U_n^3 f_3(\cdot,\theta) = o_p(1/n) \tag{24}$$

uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$. Corollary 8 will do the job.

We show in the next subsection that each class $\{f_i(\cdot,\theta) : \theta \in \Theta\}$ is Euclidean for the constant envelope 1. Equation (24) will follow from Corollary 8 provided

(i) $P^2|f_2(\cdot,\theta)| \to 0 \quad \text{as} \quad \theta \to \mathbf{0}$,

(ii) $P^3|f_3(\cdot,\theta)| \to 0 \quad \text{as} \quad \theta \to \mathbf{0}$.

We will show (ii). The proof of (i) is similar.

Recall the definition of $f(z_1, z_2, z_3, \theta)$ given above. It follows from A1 that

$$P^3\{x_1'\beta(\mathbf{0}) = x_3'\beta(\mathbf{0})\} = 0\,.$$

Deduce that $f(z_1, z_2, z_3, \cdot)$ is continuous at $\mathbf{0}$ for P^3 almost all (z_1, z_2, z_3). The proof of (6) given by Serfling (1980, pp.177-178) reveals that f_3 equals f plus or minus terms, each of which is a conditional expectation of f given zero, one, or two of its arguments. Since f is uniformly bounded in all of its arguments, a dominated convergence argument shows that $f_3(z_1, z_2, z_3, \cdot)$ is continuous at $\mathbf{0}$ for P^3 almost all (z_1, z_2, z_3). Since f_3 is also uniformly bounded in all of its arguments, another dominated convergence argument establishes (ii), which, along with (i), proves (24).

Put it all together. Combine (21), (22), (23), and (24) to get to get (19). This proves the result. □

Corollary. *If A1 through A5 hold, then*

$$\sqrt{n}(\beta_n - \beta_0) \Longrightarrow (W, 0)$$

where W *has the* $N(\mathbf{0}, V^{-1}\Delta V^{-1})$ *distribution from the last result.*

7.4. Euclidean Properties. Consider the class $\mathcal{F} = \{f(\cdot,\beta) : \beta \in \mathcal{B}\}$ where, for each (z_1, z_2, z_3) in $\mathcal{S}^3$ and each β in $\mathcal{B}$,

$$f(z_1, z_2, z_3, \beta) = \{y_1 > y_2\}\{x_1'\beta > x_3'\beta\}\,.$$

In this subsection, we show that $\mathcal{F}$ is Euclidean for the constant envelope 1. The Euclidean properties of the classes of functions encountered in the last two subsections can be deduced from this fact and Lemma 6 in Section 6. The reparametrization in terms of θ has no effect on the Euclidean properties of the classes encountered in the last subsection. We assume that the reader is familiar with the notions of a polynomial class of sets and the graph of a function as defined in Nolan and Pollard (1987).

Let t, γ, γ_1, γ_2, and γ_3 be a real numbers. Let δ_1, δ_2, and δ_3 be vectors in $\mathbb{R}^d$. For each (z_1, z_2, z_3) in $\mathcal{S}^3$, define

$$g(z_1, z_2, z_3, t\ ;\ \gamma, \{\gamma_i\}, \{\delta_i\}) = \gamma t + \sum_i \gamma_i y_i + \sum_i \delta_i' x_i$$

and

$$\mathcal{G} = \{g(\cdot,\cdot,\cdot,\cdot\ ;\ \gamma, \{\gamma_i\}, \{\delta_i\}) : \gamma, \gamma_i \in \mathbb{R},\ \delta_i \in \mathbb{R}^d,\ i = 1,2,3\}\,.$$

Notice that $\mathcal{G}$ is a $(3d+4)$-dimensional vector space of real-valued functions on $\mathcal{S}^3 \otimes \mathbb{R}$. By Lemma 18(ii) in Nolan and Pollard (1987), the class of sets of the form $\{g \geq r\}$ or $\{g > r\}$ with $g \in \mathcal{G}$ and $r \in \mathbb{R}$ is a polynomial class. We use this fact below to show that the set of graphs of functions belonging to $\mathcal{F}$ forms a polynomial class of sets. The Euclidean nature of $\mathcal{F}$ will then follow from Lemma 19 in Nolan and Pollard (1987).

For each $\beta \in \mathcal{B}$,

$$\begin{aligned}
\text{graph}(f(\cdot,\beta)) &= \{(z_1, z_2, z_3, t) \in \mathcal{S}^3 \otimes \mathbb{R} : 0 < t < f(z_1, z_2, z_3, \beta)\} \\
&= \{y_1 - y_2 > 0\}\{x_1'\beta - x_3'\beta > 0\}\{t \geq 1\}^c\{t > 0\} \\
&= \{g_1 > 0\}\{g_2 > 0\}\{g_3 \geq 1\}^c\{g_4 > 0\}
\end{aligned}$$

for $g_i \in \mathcal{G},\ i = 1,2,3,4$. The graph of $f(\cdot,\beta)$ is the intersection of four sets, three of which belong to a polynomial class, and the fourth is the complement of a set belonging to a polynomial class. Deduce from Lemma 18(i) in Nolan and Pollard (1987) that $\{\text{graph}(f) : f \in \mathcal{F}\}$ forms a (subset of a) polynomial class of sets.

Acknowledgements. I wish to thank Chris Cavanagh for sharing his ideas with me. Especially, I wish to thank my thesis advisor, David Pollard, for his generous gifts of time and insight throughout the course of this research.

REFERENCES

Arcones, M. A. and Giné, E. (1991). Limit theorems for U-processes. Preprint, Mathematical Sciences Research Institute, Berkeley, CA. and University of Connecticut, Storrs, CT.

Amemiya, T. (1985). *Advanced Econometrics.* Harvard Univ. Press, Cambridge, Mass.

Bonami, A. (1970). Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier.* **20** 335-402.

Cavanagh, C. (1990). Personal communication, Bell Communications Research, Morristown, N.J.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model. *Journal of Econometrics.* **35** 303–316.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221–233. University of California, Berkeley, CA.

Ichimura, H. (1988). Semiparametric estimation of multiple indices models: single equation estimation. Preprint, Department of Economics, University of Minnesota, Minneapolis, MN.

Klein, R. W. and Spady, R. H. (1992). An efficient semiparametric estimator for binary response models. To appear in *Econometrica.*

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Annals of Statistics* **18** 405–414.

Nolan, D. and Pollard, D. (1987). U-processes: Rates of convergence. *Annals of Statistics.* **15** 780–799.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica.* **57** 1027–1057.

de la Peña, V. H. (1990). Decoupling and Khintchine's inequalities for U-statistics. To appear in *Annals of Probability.*

Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Banach Spaces, Harmonic Analysis, and Probability Theory. Lecture Notes in Math.* **995** 123–154. Springer, New York.

Pollard D. (1982). A central limit theorem for empirical processes. *J. Austrl. Math. Soc. Ser. A.* **33** 235–248.

Pollard D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Pollard D. (1989). Unpublished lecture notes for an asymptotic methods course taught at Yale University.

Pollard D. (1990). *Empirical Processes: Theory and Applications.* Institute of Mathematical Statistics, Hayward, CA.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Sherman, R. P. (1991a). *U-processes and semi-parametric estimation.* Ph.D. thesis, Yale University.

Sherman, R. P. (1991b). Maximal inequalities for degenerate U-processes with applications to optimization estimators. Tech. Mem. TM-ARH-018-797. Bellcore, Morristown, N.J.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137.

Bellcore
445 South Street
Box 1910
Morristown, New Jersey 07962-1910