

LIMIT THEOREMS FOR THE INFINITE-DEGREE U-PROCESS

Charles Heilig and Deborah Nolan

Klemm Analysis Group, Inc. and University of California

Abstract: In this paper, we provide uniform limit theory for a U-statistic of increasing degree, also called an infinite-degree U-statistic. The stochastic process based on collections of U-statistics is referred to as a U-process, and if the U-statistic is infinite-degree, we have an infinite-degree U-process. Frees (1986) proposed a nonparametric renewal estimator which is an infinite-degree U-statistic. In a later paper, Frees (1989) provided conditions for the pointwise asymptotic theory for the infinite-degree U-statistic. To extend the pointwise results to limit theory for the infinite-degree U-process that holds uniformly over the index set, we build on existing results for U-processes of fixed degree. In particular we extend the symmetrization techniques of Nolan and Pollard (1987) and the moment inequalities of Sherman (1994) to obtain uniform weak laws of large numbers and functional central limit theory for the infinite-degree U-process.

Key words and phrases: Empirical process, functional central limit theory, U-statistic, uniform weak law of large numbers.

1. Introduction

The theory of U-statistics parallels the theory for sums of independent random variables (see Lee (1990)). In particular, Nolan and Pollard (1987) extended the analogy between U-statistics and sums of i.i.d. random variables by adapting empirical process theory to collections of U-statistics of degree 2 to obtain uniform limit theorems. Arcones and Giné (1993) and Sherman (1994) further extended these results to U-statistics of arbitrary fixed degree. In this paper, we provide uniform limit theory for U-statistics of increasing degree—infinite-degree U-statistics. To explain, consider a sequence of i.i.d. random variables, X_1, \dots, X_n , with distribution P , and a sequence of functions $\{g_m\}$, where each g_m is a real-valued kernel function of degree m with $m \leq n$, $m \rightarrow \infty$. Define the *infinite-degree U-statistic* (IDUS), $U_{n,m}(g_m) = \frac{(n-m)!}{n!} \sum g_m(X_{\pi_1}, \dots, X_{\pi_m})$, where the summation is over all subsets of m distinct indices from $1, \dots, n$. The stochastic process based on collections of U-statistics is called the U-process, and if the U-statistic is infinite-degree, we call the process an *infinite-degree U-process* (IDUP). That is, an IDUP is a collection of infinite-degree U-statistics indexed by a sequence of kernel classes: \mathcal{G}_m .

Frees (1989) considered U-statistics based on kernel functions of infinite-degree. His proposal for a nonparametric renewal estimator is an example of an IDUS (Frees (1986)). Other examples include the Nelson-Aalen cumulative hazard function under right censoring, the Kaplan-Meier survival estimator, multivariate renewal functions, and statistics based on the m out of n bootstrap. These examples are considered in Heilig (1997) and Heilig and Nolan (1998). Here we restrict ourselves to Frees' univariate nonparametric renewal estimator as an example of how to apply IDUP theory.

Consider the renewal function $N(t) = 1 + \sum_{k=1}^{\infty} P^{[k]}(-\infty, t]$, where $P^{[k]}$ denotes the k -fold convolution of P . Frees (1986) introduced a nonparametric renewal estimator which puts minimal assumptions on P . For each k , $k \leq m$, estimate $P^{[k]}$ by $U_{n,k}(f_{k,t})$, where $f_{k,t}(x_1, \dots, x_k) = \{x_1 + \dots + x_k \leq t\}$. (Note that $f_{k,t}$ is an indicator function, and we use the set itself to represent the indicator of the set.) Frees' nonparametric renewal estimator is the sum of these U-statistics, $\hat{N}_m(t) = 1 + \sum_{k=1}^m U_{n,k}(f_{k,t}) = 1 + U_{n,m}(g_{m,t})$, where $g_{m,t}(x_1, \dots, x_m) = \sum_{k=1}^m f_{k,t}(x_1, \dots, x_k)$. That is, \hat{N}_m is itself a U-statistic of degree m with a kernel function that is the sum of m kernel functions. The degree m may be as large as n , or it may grow at a slower rate to aid in the trade-off between estimator performance and computational complexity. As m grows, $\hat{N}_m(t)$ may be regarded as a U-statistic of increasing or infinite degree, and $\hat{N}_m(\cdot)$ is an infinite-degree U-process with index set $\mathcal{G}_m = \{g_{m,t}\}$. (Frees used the term "infinite-order" to describe this U-statistic, but we prefer the term "infinite-degree" in keeping with Hoeffding's original use of the term "degree.")

Frees (1989) provides conditions for the pointwise asymptotic theory for the IDUS. To extend the pointwise results to limit theory for the IDUP that holds uniformly over t , we build on existing results for U-processes of fixed degree in Nolan and Pollard (1987), Arcones and Giné (1993), and Sherman (1994). In particular we extend the symmetrization techniques of Nolan and Pollard and the moment inequalities of Sherman to obtain our results. Our results say for example, that if, for each m , the collection of kernel functions is bounded by K_m and has certain entropy properties, then provided $K_m m/n^{1/2} = o(1)$ the uniform weak law of large numbers holds. More general statements of these uniform results appear in the next section. In Section 3, we revisit the renewal example to show how these results can be applied, and the proof of the main result appears in Section 4.

2. Results

We assume the IDUS is a sequence of U-statistics of increasing degree, where the limit of the expectations of the kernels converges to some parameter of interest. We use linear functional notation to write this expectation of g_m as

$P^m(g_m) = \int g_m(x_1, \dots, x_m) dP(x_1) \cdots dP(x_m)$. In our example, the kernel g_m is constructed from a sum of m kernels, $g_m(x_1, \dots, x_m) = f_1(x_1) + f_2(x_1, x_2) + \cdots + f_m(x_1, \dots, x_m)$. We call f the subkernel and g the grand kernel.

2.1. Hoeffding decomposition

The Hoeffding decomposition for a U-statistic (Hoeffding (1948), Serfling (1980)) is a useful technique for obtaining asymptotic results. We present it here for the IDUS. It is necessary for the kernel to be symmetric in its arguments in order to apply the decomposition. If g_m is not symmetric, create a symmetric kernel as follows: $h_m(x_1, \dots, x_m) = \sum g_m(x_{\pi_1}, \dots, x_{\pi_m})/m!$. The IDUS can be written in terms of either of these kernels $U_{n,m}(h_m) = U_{n,m}(g_m)$. Throughout this paper, f denotes a subkernel, g denotes a general, possibly grand, kernel, and h a symmetric kernel.

The Hoeffding decomposition of a symmetric U-statistic is then

$$U_{n,m}(h_m) = \sum_{i=0}^m \binom{m}{i} U_{n,i}(h_{m(i)}), \quad (1)$$

where the projected kernels $h_{m(i)}$ are built up from conditioned kernels:

$$h_{m(i)}(x_1, \dots, x_i) = \sum_{k=0}^i (-1)^{i-k} \sum_{(i,k)} h_{m|k}(x_{\pi_1}, \dots, x_{\pi_k}), \quad (2)$$

$h_{m|k}(x_1, \dots, x_k) = \int h_m(x_1, \dots, x_m) dP(x_{k+1}) \cdots dP(x_m)$, and

$(i, k) = \{(\pi_1, \dots, \pi_k) \in \{1, \dots, i\}^k : \pi_j < \pi_l, \text{ for } j < l\}$.

Typically, the first-order projection of the U-statistic, $mU_{n,1}h_{m(1)}$, is the driving term behind its central limit theory. For the infinite-degree U-statistic, this term carries somewhat more delicate properties because of its dependence on m . By construction, the first-order term is an average of centered i.i.d. random variables:

$$mU_{n,1}(h_{m(1)}) = \frac{1}{n} \sum_{i=1}^n m [h_{m|1}(X_i) - P^m(h_m)]. \quad (3)$$

2.2. Euclidean case

Empirical process theory exploits the topological properties of index sets in order to effect various approximations. Suppose the index set \mathcal{G} is equipped with pseudometric d . For each $\varepsilon > 0$, the covering number $N(\varepsilon, d, \mathcal{G})$ is the smallest N for which there exist points g_1, \dots, g_N such that $\min_{i \leq N} d(g, g_i) \leq \varepsilon$ for every $g \in \mathcal{G}$.

Conditions for uniform convergence can often be stated in terms of the rate at which $N(\varepsilon, d, \mathcal{G})$ grows. While covering numbers may be difficult to discern exactly, many function classes have properties which allow their entropies to be bounded. One such property is called Euclidean (Nolan and Pollard (1987)). The class \mathcal{G} is Euclidean(A, V) for the envelope G if there exist constants A and V such that for all $\varepsilon \in (0, 1]$ and all measures Q , $N(\varepsilon, d_{Q,G}, \mathcal{G}) \leq A\varepsilon^{-V}$, where $d_{Q,G}(g, g') = [Q(|g - g'|^2)/Q(G^2)]^{1/2}$. (Note that we take $V \geq 1$ in all our applications.) Knowing that a class of functions is Euclidean aids immensely in establishing rates of uniform convergence.

We present a uniform weak law of large numbers and functional central limit theorem for the IDUP, where the collections of functions \mathcal{G}_m are Euclidean. Then follows a more general result.

Theorem 1. *Suppose the class \mathcal{G}_m is Euclidean(A, V_m) for the envelope G_m . Also suppose, G_m is bounded by K_m , and $\limsup P^m(G_m^2) < \infty$.*

- (i) *If $K_m m V_m^{1/2} = o(n^{1/2})$ and $m V_m^{1/2} = o(n^{1/2})$, then $\sup_{\mathcal{G}_m} |U_{n,m}(g) - P^m(g)| \xrightarrow{\mathbb{P}} 0$.*
- (ii) *If $K_m m^2 V_m = o(n^{1/2})$ and $m V_m^{1/2} = o(n^{1/2})$, then $\sup_{\mathcal{G}_m} \sqrt{n} |U_{n,m}(g) - P^m(g) - m U_{n,1}(h_{m(1)})| \xrightarrow{\mathbb{P}} 0$, where $h_{m(1)}$ is the first order Hoeffding projection (3).*

The proof appears in the Appendix.

2.3. General case

To present a more general result, we first introduce a pseudometric based on double samples. Let $X_1, \dots, X_n, X'_1, \dots, X'_n$ be two independent i.i.d. samples from the distribution P . Also, let $W_i = X_i$ and $W_{n+i} = X'_i$, $i = 1, \dots, n$. For $0 \leq j \leq m - 1$, define the pseudometric

$$d_j(h_j, \tilde{h}_j) = \left[\frac{U_{2n,j}(h_j - \tilde{h}_j)^2}{U_{2n,j}(H_j^2)} \right]^{1/2}, \tag{4}$$

where $U_{2n,j}$ is a U-statistic based on W_1, \dots, W_{2n} and h_j is a symmetric kernel of degree j . Note that the pseudometric depends on n ; we suppress the dependence in our notation for simplicity.

For a sequence of positive integers $\{r_j\}$, let

$$\begin{aligned} \phi(j) &= \sqrt{j!} (16r_j - 8)^{j/2} n^{-j/2} \binom{m}{j} \\ &\times \mathbb{P} \left[\left(U_{2n,j}(H_{m(j)}^2) \right)^{1/2} \int_0^{1/4} N(x, d_j, \mathcal{H}_{m(j)})^{1/2r_j} dx \right]. \end{aligned} \tag{5}$$

Theorem 2. *Suppose the class \mathcal{H}_m has P^m -square integrable envelope H_m , and $\limsup P^m(H_m^2) < \infty$. Also suppose the classes $\mathcal{H}_{m(j)}$ are totally bounded under pseudometric d_j for each j, m , and $m < n/3$.*

- (i) *If $\phi(1) + \dots + \phi(m) = o(1)$, then $\sup_{\mathcal{H}_m} |U_{n,m}(h) - P^m(h)| \xrightarrow{\mathbb{P}} 0$.*
- (ii) *If $\phi(2) + \dots + \phi(m) = o(n^{-1/2})$, then $\sup_{\mathcal{H}_m} \sqrt{n} |U_{n,m}(h) - P^m(h) - mU_{n,1}(h_{m(1)})| \xrightarrow{\mathbb{P}} 0$.*

The proof appears in Section 4.

2.4. FCLT for the Hoeffding projection

Below is a functional central limit theorem for the first-order projection of an IDUP in the case when g represents a grand kernel comprised of symmetric subkernels. In this case, it can be shown that $mh_{m(1)}(x) = \sum j [f_{j|1}(x) - P^j f_j]$, and therefore $mU_{n,1}(h_{m(1)}) = (P_n - P)(F_m)$, where $F_m(x) = \sum j f_{j|1}(x)$. It is assumed that the subkernels have a common index, say \mathcal{T} .

Theorem 3. *Suppose the following conditions hold:*

- (i) *For each $s, t \in \mathcal{T}$,*

$$\lim_{n \rightarrow \infty} \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[f_{j|1}(X_1; s) - P f_{j|1}(\cdot; s) \right] \left[f_{k|1}(X_1; t) - P f_{k|1}(\cdot; t) \right] < \infty.$$

- (ii) *The class $\{F_m\}$ is Euclidean(A, V) for the envelope \bar{F}_m , all m .*
- (iii) $\limsup_n P(\bar{F}_m(X)^2) < \infty$.
- (iv) $\limsup_n P(\bar{F}_m(X)^2 \{ \bar{F}_m > \varepsilon n^{1/2} \}) = 0$ for each $\varepsilon > 0$.
- (v) *For $\rho_n^2(s, t) = \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[f_{j|1}(X_1; s) - f_{j|1}(X_1; t) \right] \left[f_{k|1}(X_1; s) - f_{k|1}(X_1; t) \right]$, the limit $\rho(s, t) = \lim_{n \rightarrow \infty} \rho_n(s, t)$ is well-defined, and for all deterministic sequences $\{s_n\}$ and $\{t_n\}$, if $\rho(s_n, t_n) \rightarrow 0$, then $\rho_n(s_n, t_n) \rightarrow 0$. Then $n^{1/2}(P_n - P)(F_m)$ converges in distribution to a mean-zero Gaussian process which is uniformly ρ -continuous with variance/covariance kernel,*

$$\sigma(s, t) = \lim_{n \rightarrow \infty} \sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[f_{j|1}(X_1; s) - P f_{j|1}(\cdot; s) \right] \left[f_{k|1}(X_1; t) - P f_{k|1}(\cdot; t) \right].$$

To prove this result, invoke Theorem 10.6 of Pollard (1990).

2.5. Discussion

We have presented two sets of conditions for proving uniform weak laws of large numbers for infinite-degree U-processes. The main difference between the conditions relates to the properties of the kernel classes. Theorem 1 is a

special case of Theorem 2 when \mathcal{G}_m is Euclidean. According to Theorem 1, if \mathcal{G}_m is Euclidean(A, V) for a constant envelope then if $m = o(n^{1/2})$ a uniform law of large numbers follows, and if $m = o(n^{1/4})$ a uniform central limit theorem follows.

The constraints on the rate at which m may grow are due to the Hoeffding decomposition. In the more general case of Theorem 2 these constraints appear as restrictions on the sum of the $\phi(j)$ in (5). When m is fixed, we recover results comparable to those of Sherman (1994). These results are not as strong as those of Arcones and Giné (1993) for the finite case because they employ moment inequalities rather than exponential inequalities. We are unable to extend Arcones and Giné's results to the infinite degree case because they rely on a symmetrization inequality of de la Peña's (1992) which incurs upper bounds that grow with m much too quickly.

Several classes of statistics share commonalities with infinite-degree U-statistics. We mention a few, including infinite-degree V-statistics, partial-sum U-processes, symmetric statistics and elementary symmetric polynomials. Blom (1976) introduced the incomplete U-statistic, which is an average over a subset of all $(n)_k$ k -tuples. While these statistics are still unbiased, they are no longer minimum-variance unbiased. Nonetheless, for some subsets of k -tuples, the loss in efficiency may be minimized and may even vanish asymptotically. Frees (1989) considers a random subsampling scheme for an incomplete IDUS. Recently, Shieh (1994) defines infinite-degree V-statistics by analogy with the IDUS. While V-statistics are not generally unbiased, they may enjoy more efficient computation algorithms than U-statistics. Politis and Romano (1994) and Bickel, Götze and van Zwet (1997) consider statistics based on resampling procedures. Among the procedures summarized in Bickel, Götze and van Zwet (1997), the m out of n bootstrap is an infinite-degree V-statistic and the n choose m bootstrap is an infinite-degree U-statistic. Grübel and Pitts (1993) and Harel, O'Connell and Schneider (1995) study an infinite-degree V-process estimator of the renewal function; see the next section for more details. Kohatsu-Higa (1991) defines the partial-sum U-process, and obtains results related to those of Dynkin and Mandelbaum (1983) for symmetric statistics. Heilig (1997) lists several further references regarding symmetric statistics, including weak convergence, invariance principles, and Berry-Esseen bounds.

Regarding measurability, in our applications suprema can be taken over a countable class of events. However, the contributing theory can be generalized to uncountable classes. For a discussion of these measurability issues, we refer to Chapter 10 in Dudley (1984) and Appendix C in Pollard (1984).

3. Example

Recall from the Introduction that Frees proposed $\hat{N}_m(t)$, a U-statistic of degree m , as an estimator for the renewal function. This U-statistic arose as a sum of U-statistics with kernels of degree $k = 1, \dots, m$, and although each of the subkernels $f_{k,t} = \{x_1 + \dots + x_k \leq t\}$ is symmetric, the grand kernel $g_{m,t} = \sum_{k=1}^m f_{k,t}$ is not symmetric. Also, although the renewal estimator is biased for each m , its expectation converges to the renewal function N , as $m \rightarrow \infty$.

Lemma 4. *Let P be a probability measure with positive mean, and let τ be a finite constant. Then for any $m \leq n$ such that $m \rightarrow \infty$, $\sup_{t \in [0, \tau]} |\hat{N}_m(t) - N(t)| \xrightarrow{\mathbb{P}} 0$.*

Lemma 5. *Suppose X is a random variable with probability measure P . Assume that P has positive mean, finite variance, and that for some $\eta > 0$, $\int [X^-]^{5+\eta} dP < \infty$. Let τ be a finite constant. Then for any $m \geq n^{1/(6+2\eta)}$, the process $n^{1/2}(\hat{N}_m - N)$ obeys a functional central limit theorem over $[0, \tau]$, where the finite dimensional distributions are Gaussian with variance/covariance matrix determined by $\sum_{j,k}^{\infty} jk [\mathbb{P}[P^{[j-1]}(s - X)P^{[k-1]}(t - X)] - P^{[j]}(s)P^{[k]}(t)]$.*

The proofs are found in the Appendix.

Comments. In contrast to our approach, Grübel and Pitts (1993) study the infinite-degree V-process. Their estimator is the nonparametric maximum likelihood estimator for this model. They define a class of metrics that enable them to get uniform convergence and a limiting Gaussian process over the entire real line, using linearization and the continuous mapping theorem. They assume that, $\eta > 0$, $\mathbb{P}|X|^{2+\eta} < \infty$ for the uniform strong law of large numbers, and $\mathbb{P}|X|^{4+\eta} < \infty$ for the functional central limit theorem. Harel, O’Cinneide and Schneider (1995) study the same estimator, restricted to non-negative random variables. Schneider, Lin and O’Cinneide (1990) perform computational comparisons of the two estimators applied to non-negative random variables. They find that the IDUP has a smaller mean squared error, but the infinite degree V-process is computationally less burdensome. Grübel and Pitts (1993) suggest that Frees’ estimator may be more suitable when the sample size is small, one is interested in relatively small values of t , or the renewal times are positive.

4. Proof of the General Case

We present first two inequalities that are required for the proof of Theorem 2. The moment inequality is found in Bonami (1970), and the chaining inequality, adapted from Pisier (1983), can be found in Nolan and Pollard (1987).

Lemma 6. (Moment inequality) *For every integer $s \geq 1$, we have*

$$\mathbb{P} \left[\sum_{\pi \in (n,k)} \sigma_{\pi_1} \cdots \sigma_{\pi_k} u(\pi_1, \dots, \pi_k) \right]^{2s} \leq (2s - 1)^{ks} \left[\sum_{\pi \in (n,k)} u(\pi_1, \dots, \pi_k)^2 \right]^s.$$

Lemma 7. (Chaining inequality) *Let Ψ be a convex, strictly increasing function on $[0, \infty)$ with $0 \leq \Psi(0) \leq 1$. Let p be a positive integer. Suppose the set T is endowed with pseudometric d , and the stochastic process $\{Z(t) : t \in T\}$ satisfies the following conditions.*

- (i) *If $d(s, t) = 0$, then $Z(s) = Z(t)$ almost surely.*
- (ii) *If $d(s, t) > 0$, then $\mathbb{P}[\Psi(|Z(s) - Z(t)|^p/d(s, t)^p)] \leq 1$.*
- (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

$$\left[\mathbb{P}(\sup_T |Z(t) - Z(t_0)|^p) \right]^{1/p} \leq 8 \int_0^{\delta/4} \left[\Psi^{-1}(N(x, d, T)) \right]^{1/p} dx,$$

where $N(x, d, T)$ is the covering number defined in the previous section.

To prove our results, we make use of the Hoeffding decomposition (1). We extend Sherman’s maximal inequality to degenerate U-processes whose degree may be arbitrarily large by paying close attention to all constants that depend on the kernel. We introduce products of j sign variables for a degenerate kernel of degree j , and we obtain our maximal inequality via chaining and moment inequalities applied to each component of the Hoeffding decomposition.

Proof of Theorem 2. To begin, consider a class of kernels \mathcal{H} , degenerate and of degree j . For an integer $b = 0, \dots, 2^j - 1$, let (b_1, \dots, b_j) be its binary expansion, and for each $h \in \mathcal{H}$, define for $\pi \in (n)_j$,

$$h^\circ(\mathbf{X}_\pi, \mathbf{X}'_\pi) = \sum_{b=0}^{2^j-1} (-1)^{\sum b_i} h[b_1 X_{\pi_1} + (1 - b_1) X'_{\pi_1}, \dots, b_j X_{\pi_j} + (1 - b_j) X'_{\pi_j}].$$

Since h is degenerate, we have $\mathbb{P}[h^\circ(\mathbf{X}_\pi, \mathbf{X}'_\pi) | X_1, \dots, X_n] = h(X_{\pi_1}, \dots, X_{\pi_j})$. Use this fact and Jensen’s inequality (applied conditionally) to obtain the upper bound, for $p \geq 1$,

$$\mathbb{P}(\sup_{\mathcal{H}} |U_{n,j} h|^p) \leq \mathbb{P}(\sup_{\mathcal{H}} \left| \frac{(n-j)!}{n!} \sum_{\pi \in (n)_j} h^\circ(\mathbf{X}_\pi, \mathbf{X}'_\pi) \right|^p),$$

where $(n)_m = \{(\pi_1, \dots, \pi_m) \in \{1, \dots, n\}^m : \pi_j \neq \pi_k, \text{ for } j \neq k\}$. It also follows that $\mathbb{P}(\sup_{\mathcal{H}} |U_{n,j}(h)|^p) \leq \mathbb{P}(\sup_{\mathcal{H}} |U_{n,j}^\circ(h)|^p)$, where $U_{n,j}^\circ(h) = (n-j)!/n! \sum \sigma_{\pi_1} \dots \sigma_{\pi_j} h^\circ(\mathbf{X}_\pi, \mathbf{X}'_\pi)$. This inequality allows us to consider the completely sign-symmetrized process $U_{n,j}^\circ$ in place of the original degenerate process $U_{n,j}$.

We now extend Sherman’s maximal inequality (1994, Section 3) to infinite-degree U-processes.

Theorem 8. (Maximal inequality) *Let \mathcal{H} be a class of degenerate functions $\{h\}$ of j arguments with envelope H . Assume $P^j H^2 < \infty$. Let p and r be positive integers, and let $\Gamma_j = [64j!(16pr - 8)^j]^{p/2}$, $\tau_{n,j} = [U_{2n,j}(H^2)]^{1/2}$, and*

$$\delta_{n,j} = \sup_{\mathcal{H}} [U_{2n,j}(h^2)]^{1/2} / 4 [U_{2n,j}(H^2)]^{1/2}.$$

Then

$$\mathbb{P}[\sup_{\mathcal{H}} |n^{j/2} U_{n,j}(h)|^p] \leq \Gamma_j \mathbb{P} \left[\tau_{n,j} \int_0^{\delta_{n,j}} N(x, d_j, \mathcal{H})^{1/2pr} dx \right]^p.$$

Proof. We apply the chaining inequality (Lemma 7) to a normalized version of the sign-symmetrized U-process $U_{n,j}^\circ$. In Lemma 7, let the stochastic process $Z(t)$ be $n^{j/2} U_{n,j}^\circ(h) / \tau_{n,j}$; let the pseudometric be $d_j(h, h')$ defined in (4); and let the convex function $\Psi(x)$ be x^{2r} / γ . The factor γ will be chosen to satisfy the second condition of the chaining lemma, which we establish now. Condition on the double sample \mathcal{W} to find the bound

$$\begin{aligned} & \mathbb{P}_{\mathcal{W}} \left[\frac{n^{j/2} U_{n,j}^\circ(h - h')}{[U_{2n,j}(h - h')^2]^{1/2}} \right]^{2pr} \\ & \leq 4^{jpr} \mathbb{P}_{\mathcal{W}} \left[\sum_{\pi \in (n)_j} \sigma_{\pi_1} \cdots \sigma_{\pi_j} \frac{(h^\circ - h'^\circ)(\mathbf{X}_\pi, \mathbf{X}'_\pi)}{[\sum_{i \in (2n)_j} (h - h')(W_{i_1}, \dots, W_{i_j})^2]^{1/2}} \right]^{2pr} \\ & \leq 4^{jpr} [j!^{pr} (2pr - 1)^{jpr}] \mathbb{P}_{\mathcal{W}} \left[\frac{\sum_{\pi \in (n)_j} (h^\circ - h'^\circ)(\mathbf{X}_\pi, \mathbf{X}'_\pi)^2}{[\sum_{i \in (2n)_j} (h - h')(W_{i_1}, \dots, W_{i_j})^2]} \right]^{pr} \\ & \leq [j!(16pr - 8)^j]^{pr}. \end{aligned}$$

The first inequality follows from: $n^{j/2}(n - j)! \sqrt{(2n)!} / [n! \sqrt{(2n - j)!}] < 2^j$, for $j < n/3$. The middle inequality is due to Bonami's inequality (Lemma 6). The final inequality follows from the Cauchy-Schwarz inequality, which implies $\sum [h^\circ(\mathbf{X}_\pi, \mathbf{X}'_\pi)]^2 \leq 2^j \sum h(W_{i_1}, \dots, W_{i_j})^2$. Letting $\gamma = [j!(16pr - 8)^j]^{pr}$, the second condition of the chaining lemma is met. Take t_0 to be the zero function to verify the third condition. For the continuity condition, we have

$$|U_{n,j}^\circ(h - h')| \leq \left[\frac{2^j (2n)_j}{(n)_j} U_{2n,j}(h - h')^2 \right]^{1/2} < 2^j d_j(h, h') \tau_{n,j}.$$

This inequality also gives us the first condition. With $\Psi^{-1}(y) = (\gamma y)^{1/2r}$, the chaining lemma gives

$$\mathbb{P}_{\mathcal{W}}(\sup_{\mathcal{H}} |n^{j/2} U_{n,j}^\circ h / \tau_{n,j}|^p) \leq \left[8 \int_0^{\delta_{n,j}} [\gamma N(x, d_j, \mathcal{H})]^{1/2pr} dx \right]^p.$$

Multiply through by $|\tau_{n,j}|^p$, and take expectations to get the desired result.

The maximal inequality is the centerpiece of both the uniform weak law of large numbers and the functional central limit theorem. Apply this inequality to terms of the Hoeffding decomposition of the infinite-degree U-process, each of which is a completely degenerate U-process. To begin, consider a sequence of kernel classes \mathcal{H}_m of an infinite-degree U-process, with associated envelopes H_m where $P^m H_m < \infty$. For $j = 1, \dots, m$, let r_j be a sequence of integers, and $\mathcal{H}_{m(j)}$ be the function classes associated with the kernel projections, each with envelope $H_{m(j)}$. Recall that each of the projected kernels is a sum of conditioned kernels (2). If each conditioned kernel is bounded by some function, say $|h_{m|k}| \leq H_{m|k}$ (a convenient abuse of the conditioning notation), then for every $h \in \mathcal{H}$,

$$|h_{m(j)}(x_1, \dots, x_j)| \leq H_{m(j)}(x_1, \dots, x_j) \stackrel{\text{def}}{=} \sum_{k=0}^j \sum_{\pi \in (j,k)} H_{m|k}(x_{\pi_1}, \dots, x_{\pi_k}).$$

Then, by Theorem 8, bound the j th Hoeffding projection

$$\mathbb{P} \left[\sup_{\mathcal{H}_{m(j)}} \left| \binom{m}{j} U_{n,j}(h_{m(j)}) \right| \right] \leq 8\phi(j), \tag{6}$$

where $\phi(j)$ is defined in (5). Note the upper limit of $1/4$ for the integral is valid, since $\delta_{n,j} \leq 1/4$. Sum (6) over j , and the conditions in (i) and (ii) imply the result.

Acknowledgement

The authors are grateful to the referees for making many valuable comments, including pointing out an error in the original proof.

Appendix

For the proofs of Lemmas 4 and 5 we first need to establish the following result.

Lemma 9. *Suppose \mathcal{G}_m is Euclidean(A_m, V_m) for the envelope G_m . Then the following hold*

- (i) \mathcal{H}_m is Euclidean (A_m, V_m) for the envelope $H_m = \sum G_m(x_{\pi_1}, \dots, x_{\pi_m})/m!$
- (ii) If G_m is bounded above by a constant, say K_m , then $\mathcal{H}_{m|j}$ is Euclidean (A_m, V_m) for the envelope $H_{m|j} = K_m$.
- (iii) If G_m is bounded above by K_m , then $\mathcal{H}_{m(j)}$ is Euclidean ($A_m, 2V_m$) for the envelope $H_{m(j)} = 2^j K_m$.

Proof. By the Euclidean property of \mathcal{G} , for any $g \in \mathcal{G}$ there exists a g^* in the approximating class such that for all Q , $\int |g - g^*|^2 dQ \leq \varepsilon^2 \int G^2 dQ$. Then for $h = \sum g(x_{\pi_1}, \dots, x_{\pi_m})/m!$, approximate it by $h^* = \sum g^*(x_{\pi_1}, \dots, x_{\pi_m})/m!$. It follows from Jensen's inequality that

$$\int |h - h^*|^2 dQ \leq \varepsilon^2 \int \sum_{\pi \in (m)_m} G(x_{\pi_1}, \dots, x_{\pi_m})^2 dQ(x_1, \dots, x_m)/m!.$$

Result (i) is established. Statement (ii) follows from Corollary 21 of Nolan and Pollard (1987), which states that for \mathcal{F} , a uniformly bounded Euclidean class of functions on $\mathcal{X} \times \mathcal{X}$, the class of functions $\{\int f(x, y)dQ(x) : f \in \mathcal{F}\}$ is Euclidean for the same constants and envelope.

Statement (iii) follows from splitting $\mathcal{H}_{m(j)}$ into two parts, one for the positive sums and one for the negative sums, i.e., according as the exponent of $(-1)^{j-k}$ is even or not. Each of these collections of functions is Euclidean (A_m, V_m) for the envelope $2^{(j-1)}K_m$. Corollary 17 of Nolan and Pollard (1987) says that if \mathcal{F} is Euclidean (A_1, V_1) for the envelope F and if \mathcal{G} is Euclidean (A_2, V_2) for the envelope G , then $\mathcal{F} + \mathcal{G} = \{f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is Euclidean $(A_1 A_2, V_1 + V_2)$ for the envelope $F + G$.

Proof of Lemma 4. A symmetric version of the renewal estimator is: $1 + U_{n,m}(h_{m,t})$, where $h_{m,t}(x_1, \dots, x_m) = \sum_{\pi \in (m)_m} g_{m,t}(x_{\pi_1}, \dots, x_{\pi_m})/m!$. Consider the entropy property of this index class. Define the graph $\Gamma(f)$ for a real-valued function f on \mathcal{S}^k : $\Gamma(f) = \{(\mathbf{x}, y) \in \mathcal{S}^k \times \mathbb{R} : 0 \leq y \leq f(\mathbf{x}) \text{ or } f(\mathbf{x}) \leq y \leq 0\}$. The entropy of the index classes relies on one important property of the graphs of the subkernels: if $s < t$ then $\Gamma(f_{k,s}) \subseteq \Gamma(f_{k,t})$. This property implies that $\{f_{k,t}\}$ is Euclidean $(A, 2)$ for the envelope 1 and some constant A . The entropy of $\{f_{k,t}\}$ is determined by the ordering property alone. Since this ordering property carries over to the graphs of $g \in \mathcal{G}_m$, they too are Euclidean $(A, 2)$ for the constant envelope m . Results such as these can be found in Dudley (1985, p.496 and Example 5.4, p.506). It then follows from Lemma 9 that \mathcal{H}_m is Euclidean $(A, 2)$ for the envelope m , and that $\mathcal{H}_{m(i)}$ is Euclidean $(A, 4)$ for the envelope $2^i m$.

Next prepare to apply Theorem 1. Replace N by $\mathbb{P}(\hat{N}_m)$. This is possible because the condition that $m \rightarrow \infty$ implies

$$\sup_{t \in [0, \tau]} |\mathbb{P}[\hat{N}_m(t)] - N(t)| = \sum_{k > m} P^{[k]}(-\infty, \tau) \rightarrow 0.$$

Next we need only consider $m \leq m^*$, where $m^* = \log n/4$. For if $m \geq m^*$,

$$\begin{aligned} \mathbb{P}[\sup_{t \in [0, \tau]} |\hat{N}_m(t) - \hat{N}_{m^*}(t)|] &= \sum_{k=m^*+1}^m \mathbb{P}[U_{n,k}(f_{k,\tau})] \\ &= \sum_{k=m^*+1}^m P^{[k]}(-\infty, \tau) \rightarrow 0. \end{aligned}$$

Now apply Theorem 1 to the IDUP $\hat{N}_m(t) - \mathbb{P}[\hat{N}_m(t)]$, i.e. $U_{n,m}(g_{m,t}) - P^m(g_{m,t})$. The IDUP is indexed by the grand kernel \mathcal{G}_m , which has been shown to be Euclidean($A, 2$) for the envelope m . The conditions of the theorem are easily met.

Proof of Lemma 5. As in the proof of Lemma 4, we replace N by $\mathbb{P}(\hat{N}_m)$. To make this replacement, we use the following result (Gut (1974, Theorem 2.1)): for $r > 2$, $\mathbb{P}[X^-]^r < \infty$ implies $\sum k^{r-2}P^{[k]}(t) < \infty$. Therefore, for $m \geq m^* = n^{1/(2r-4)}$,

$$n^{1/2} \sup_{t \in [0, \tau]} \left| \sum_{k > m} P^{[k]}(t) \right| \leq \sum_{k > m+1} k^{r-2} P^{[k]}(\tau) \rightarrow 0,$$

$$n^{1/2} \mathbb{P} \left[\sup_{t \in [0, \tau]} \left| \hat{N}_m(t) - \hat{N}_{m^*}(t) \right| \right] = n^{1/2} \sum_{k=m^*+1}^m P^{[k]}(\tau) \rightarrow 0.$$

Now we can work with the IDUP $\hat{N}_m(t) - \mathbb{P}[\hat{N}_m(t)] = U_{n,m}(g_{m,t}) - P^m(g_{m,t})$, for $m \leq m^*$. Theorem 1 holds, for $m = o(n^{1/6})$. When $r > 5$, this constraint on m^* is satisfied.

Finally, for the functional central limit theorem, apply Theorem 3 to the first-order projection $n^{1/2}(P_n - P)F_{m,t}$, where $F_{m,t}(x) = \sum j P^{[j-1]}(-\infty, t - x]$. Use the following bound to verify the first and fifth conditions of the Theorem,

$$\sum_{j=1}^m \sum_{k=1}^m jk \mathbb{P} \left[P^{[j-1]}(-\infty, s - X_1] P^{[k-1]}(-\infty, t - X_1] - P^{[j]}(-\infty, s] P^{[k]}(-\infty, t] \right]$$

$$\leq \sum_{j=1}^m \sum_{k=1}^m jk P^{[j]}(s) \wedge P^{[k]}(t).$$

In addition let $F_{m,\tau}$ be the envelope for $\{F_{m,t}\}$. Since $\{f_{k,t}\}$ is Euclidean($A, 1$) for envelope 1, the collection $\{F_{m,t}\}$ is Euclidean($A, 1$) and $P^m(F_{m,\tau}^2) < \infty$. All the conditions of Theorem 3 are met. The result now follows.

Proof of Theorem 1. The proof follows directly from Theorem 2. Consider $\phi(j)$ in (5). Take $r_j = V_m + 1$ and bound the integral in (5) by a constant. Also, from Lemma 9 we have a bound of $2^j K_m$ on $H_{m(j)}$. Put these two bounds together to find, $j = 1, \dots, m$,

$$\phi(j) \leq CK_m n^{-j/2} \binom{m}{j} (j!)^{1/2} 32^{j/2} (2V_m + 1)^{(j/2)} = \psi(j).$$

Note that for $j = 2, \dots, m$

$$\psi(j) = \psi(j - 1) \sqrt{32} \sqrt{2V_m + 1} \frac{(m - j + 1)}{(jn)^{1/2}}$$

$$\begin{aligned} &\leq \psi(j-1) \frac{12m\sqrt{V_m}}{n^{1/2}} \\ &\leq \psi(1)(12m\sqrt{V_m}/n^{1/2})^{j-1}. \end{aligned}$$

Therefore, for some constant C , $\phi(1) + \dots + \phi(m) \leq CK_m m V_m^{1/2} n^{-1/2}$ and $\phi(2) + \dots + \phi(m) \leq CK_m m^2 V_m n^{-1}$. The results now follow from the conditions imposed on m , K_m , and V_m .

References

- Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21**, 1494-1542.
- Bickel, P. J., Götze, F. and van Zwet, W. R. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statist. Sinica* **7**, 1-31.
- Blom, G. (1976). Some properties of incomplete U-statistics. *Biometrika* **63**, 573-580.
- Bonami, A. (1970). Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier* **20**, 335-402.
- de la Peña, V. H. (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.* **20**, 1877-1892.
- Dudley, R. M. (1984). A course on empirical processes. *École D'été de Probabilités de Saint-Flour, XII—1982*. Vol. 1097 of *Lecture Notes in Math.*, 1-142. Springer, New York.
- Dudley, R. M. (1985). The structure of some Vapnik-Červonenkis classes. *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer II*, 495-508. Wadsworth, Belmont, CA.
- Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Ann. Statist.* **11**, 739-745.
- Frees, E. W. (1986). Nonparametric renewal function estimation. *Ann. Statist.* **14**, 1366-1378.
- Frees, E. W. (1989). Infinite order U-statistics. *Scand. J. Statist.* **16**, 29-45.
- Grübel, R. and Pitts, S. M. (1993). Nonparametric estimation in renewal theory I: the empirical renewal function. *Ann. Statist.* **21**, 1431-1451.
- Gut, A. (1974). On the moments and limit distributions of some first passage times. *Ann. Probab.* **2**, 277-308.
- Harel, M., O'Connell, C. A. and Schneider, H. (1995). Asymptotics of the sample renewal function. *J. Math. Anal. Appl.* **189**, 240-255.
- Heilig, C. M. (1997). *An Empirical Process Approach to U-processes of Increasing Degree*. Ph. D. dissertation, University of California, Berkeley.
- Heilig, C. M. and Nolan, D. A. (1998). Applications of the infinite-degree U-process. Manuscript under review.
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.
- Kohatsu-Higa, A. (1991). Weak convergence of infinite order U-processes. *Statist. Probab. Lett.* **12**, 145-150.
- Lee, A. J. (1990). *U-statistics: Theory and Practice*. Marcel Dekker, New York.
- Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15**, 780-799.
- Nolan, D. and Pollard, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16**, 1291-1298.
- Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Banach Spaces, Harmonic Analysis, and Probability Theory: Proceedings of the Special Year in Analysis, Lecture Notes in Math* **995**, 123-154. Springer-Verlag, New York.

- Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031-2050.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.
- Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Vol. 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. The Institute of Mathematical Statistics and the American Statistical Association, Hayward, CA; Alexandria, VA.
- Schneider, H., Lin, B. S. and O'Kinneide, C. (1990). Comparison of nonparametric estimators for the renewal function. *J. Roy. Statist. Soc. Ser. C* **39**, 55-61.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.
- Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22**, 439-459.
- Shieh, G. S. (1994). Infinite order V-statistics. *Statist. Probab. Lett.* **20**, 75-80.

Klemm Analysis Group, Inc., 2200 Century Parkway, Ste 725, Atlanta GA 30345, U.S.A.

E-mail: cmheilig@klemmanalysis.com

University of California, 367 Evans Hall # 3860, Berkeley CA 94720-3860, U.S.A.

E-mail: nolan@stat.berkeley.edu

(Received March 1998; accepted June 2000)