Limit Theorems for the Infinite-Degree U-process

Charles Heilig and Deborah Nolan

Klemm Analysis Group, Inc., and University of California at Berkeley

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 non-parametric 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), de la Peña (1992), and Arcones and Giné (1993), 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.

Keywords: U-statistic, empirical process, functional central limit theory, uniform weak law of large numbers.

1 Introduction

The theory of U-statistics parallels the theory for sums of independent random variables. A survey of laws of large numbers and central limit theorems for U-statistics can be found in Lee (1990). 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) extended these results to U-statistics of arbitrary fixed degree. In this paper, we provide uniform limit theory for a U-statistic of increasing degree—an infinite-degree U-statistic. For a sequence of i.i.d. random variables, $X_1, \dots X_n$, with distribution P and a sequence of functions $\{g_m\}$, where g_m is a real-valued kernel of degree $m(n) \leq n$, define the *infinite-degree U-statistic* (IDUS),

$$U_{n,m} g_m = \frac{(n-m)!}{n!} \sum_{\pi \in (n)_m} g_m(X_{\pi_1}, \dots, X_{\pi_m}),$$

where

$$(n)_m = \{(\pi_1, \dots, \pi_m) \in \{1, \dots, n\}^m : \pi_j \neq \pi_k, \text{ for } j \neq k\}.$$

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:

$$\{U_{n,m} g_m : \ g_m \in \mathcal{G}_m\}.$$

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.

In renewal theory and survival analysis, the support of P is generally assumed to be the positive real line; in sequential analysis, its support may extend to negative real numbers, but its mean is assumed positive. The number of renewals by time t is

$$1 + \sup\{n : \sum_{i=1}^{n} X_i \leq t\}.$$

Its expectation is the renewal function, $N(t)$. Denote the distribution of $\sum_{i=1}^{k} X_i$, the k-fold convolution of P, by $P^{[k]}$. The renewal function can be written as follows:

$$N(t) = 1 + \sum_{k=1}^{\infty} P^{[k]}(-\infty, t].$$

Frees (1986) introduced a nonparametric renewal estimator which makes minimal assumptions about P. For each k, $k \leq m(n)$, construct an optimal, unbiased estimator for $P^{[k]}$ by drawing on the theory of U-statistics: the minimum-variance unbiased estimator of $P^{[k]}(-\infty, t]$ is in fact the average of indicators over the entire sample is

$$P_n^{[k]}(-\infty, t] = \binom{n}{k}^{-1} \sum_{\pi \in (n,k)} \{X_{\pi_1} + \cdots + X_{\pi_k} \leq t\},$$

where

$$(n, k) = \left\{(\pi_1, \dots, \pi_k) \in \{1, \dots, n\}^k : \pi_i < \pi_j, \text{ for } i < j\right\}.$$

The estimator is a U-statistic of degree k with kernel function being the indicator function,

$$\{x_1 + \cdots + x_k \leq t\}.$$

We use the set itself to represent the indicator of the set. Frees' nonparametric renewal estimator is the sum of these U-statistics:

$$\begin{aligned}\hat{N}_m(t) &= 1 + \sum_{k=1}^{m(n)} P_n^{[k]}(-\infty, t] \\ &= 1 + \frac{(n-m)!}{n!} \sum_{\pi \in (n)_m} \sum_{k=1}^{m} \{X_{\pi_1} + \cdots + X_{\pi_k} \le t\}.\end{aligned}$$

The second equation shows that $\hat{N}_m$ is itself a U-statistic of degree m. The design parameter $m = m(n)$ grows with n; it may be n or it may grow at a slower rate to aid in the trade-off between estimator performance and computational complexity. As $m(n)$ grows, $\hat{N}_m$ may be regarded as a U-statistic of increasing or infinite degree. (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 the index set, 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 Arcones and Giné, 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 (e.g., Euclidean with constants that do not depend on m) then provided

$$K_m m \log m / n^{1/2} = o(1),$$

the uniform weak law of large numbers holds. More general statements of these uniform results appear in the next section. The remainder of the paper is organized as follows. Notation and the main theorems appear in Section 2; in Section 3 we revisit the renewal example to show how the theorems can be applied; preliminaries for the proofs appear in Section 4; and, the proofs of the main results appear in Sections 5 and 6.

2 Notation

As in our example, 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 write the expectation of g_m as

$$P^m g_m = (P \times \cdots \times P) g_m = \int g_m(x_1, \ldots, x_m) dP(x_1) \cdots dP(x_m).$$

That is, we use linear functional notation to represent expectation. Often, as in our example, the kernel g_m is constructed from a sum of m kernels,

$$g_m(x_1, \ldots, x_m) = f_1(x_1) + f_2(x_1, x_2) + \cdots + f_m(x_1, \ldots, x_m).$$

When this happens, we call f the subkernel and g the grand kernel.

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 the permutation–symmetric kernel h_m as follows:

$$\begin{aligned} h_m(x_1, \dots, x_m) &= \frac{1}{m!} \sum_{\pi \in (m)_m} g_m(x_{\pi_1}, \dots, x_{\pi_m}) \\ &= \sum_{k=1}^{m} (m)_k^{-1} \sum_{\pi \in (m)_k} f_k(x_{\pi_1}, \dots, x_{\pi_k}). \end{aligned}$$

The IDUS can be written in terms of any of these kernels

$$U_{n,m} h_m = U_{n,m} g_m = \sum_{k=1}^{m} U_{n,k} f_k.$$

Throughout this paper, f denotes a subkernel, g denotes a general kernel, and h a symmetric kernel.

Next we define the conditioned-kernel $h_{m|k}$, a kernel of degree k created from h_m by conditioning on subsets of k arguments,

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

As the kernel h_m is symmetric, the conditioned kernel $h_{m|k}$ is necessarily symmetric. The Hoeffding decomposition is then

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

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

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

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 $m(n)$. By construction, the first-order term is an average of centered i.i.d. random variables:

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

We can see that each kernel sequence $\mathcal{G}_m$ is associated with several other kernel sequences, including the subkernels $\mathcal{F}_k$, the symmetrized kernels $\mathcal{H}_m = \{h_m\}$, the conditioned kernels $\mathcal{H}_{m|i} = \{h_{m|i}\}$, and the projected kernels $\mathcal{H}_{m(i)} = \{h_{m(i)}\}$. For kernel class $\mathcal{G}_m$, let G_m represent an envelope for $\mathcal{G}_m$; that is, $|g| \le G_m$ for every $g \in \mathcal{G}_m$. Similarly for the kernel classes, $\mathcal{H}_m$, $\mathcal{H}_{m|i}$, and $\mathcal{H}_{m(i)}$, ignore the slight abuse in notation in representing the envelopes by H_m, $H_{m|i}$, and $H_{m(i)}$, respectively.

Geometric properties of the function class 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 $\min_{i \le N} d(g, g_i) \le \varepsilon$ for every $g \in \mathcal{G}$. The covering number is finite if and only if $\mathcal{G}$ is totally bounded under d.

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,

$$N(\varepsilon, d_{Q,G}, \mathcal{G}) \leq A\varepsilon^{-V}, \tag{3}$$

where

$$d_{Q,G}(g, g') = Q|g - g'|/QG.$$

Knowing that a class of functions is Euclidean aids immensely in establishing rates of uniform convergence. In the next section, we see some special Euclidean classes of IDUP kernels. A variety of pseudometrics will be employed to establish the asymptotic properties of the IDUP. We present them here, followed by a statement of the main results. They each use empirical measures based on double and triple samples. Let $X_1, \ldots, X_n$, $\tilde{X}_1, \ldots, \tilde{X}_n$, $X'_1, \ldots, X'_n$ be three independent i.i.d. samples from the distribution P. Also, let $W_i = X_i$ and $W_{n+i} = X'_i$, $i = 1, \ldots, n$. For simplicity, we introduce the additional notation,

$$\begin{aligned}
h_m(\pi, 0) &= h(X'_{\pi_1}, \tilde{X}_{\pi_2}, \ldots, \tilde{X}_{\pi_m}) \\
h_m(\pi, m-1) &= h(X'_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_m}) \\
h_m(\pi, j) &= h(X'_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_{j+1}}, \tilde{X}_{\pi_{j+2}}, \ldots, \tilde{X}_{\pi_m}), \quad j = 1, \ldots, m-2.
\end{aligned} \tag{4}$$

For j between 0 and $m(n) - 1$, define pseudometric $d_{1,j}$ as

$$d_{1,j}(h, h') = \sum_{k=1}^{n} \left| \sum_{\pi \in (n)_m : \pi_1 = k} h(\pi, j) - h'(\pi, j) \right| / \sum_{i \in (n)_m} H(i, j).$$

For the pseudometric $d_{2,j}$ we average over the double sample via the U-process $U_{2n,j}$,

$$\begin{aligned}
U_{2n,j} h_j &= \frac{1}{(2n)_j} \sum_{pi \in (2n)_j} h_j(W_{\pi_1}, \ldots, W_{\pi_j}), \\
d_{2,j}(h_j, h'_j) &= \left[\frac{U_{2n,j}\left(h_j - h'_j\right)^2}{U_{2n,j} H_j^2} \right]^{1/2}.
\end{aligned}$$

Note that both pseudometrics depend on n. We suppress the dependence in our notation for simplicity. Also note that if $\mathcal{H}$ is Euclidean(A, V) for the envelope H_m then, for $j = 0, \ldots, m-1$,

$$N(\varepsilon, d_{1,j}, \mathcal{H}) \leq A\varepsilon^{V}.$$

A similar bound holds for covering numbers based on $d_{2,j}$.

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).

Main Results We present a uniform weak law of large numbers and functional central limit theorem for the IDUP. For the law of large numbers, two versions based on two different techniques of proof are provided. The first technique extends the symmetrization and exponential inequalities found in Nolan and Pollard (1987) and Arcones and Giné (1993). Arcones and Giné symmetrize using m samples, where m is finite, and they apply decoupling inequalities developed by de la Peña (1992). However, the constants in de la Peña's inequalities grow much too quickly to be used with U-processes of increasing degree. We use a different symmetrization based on three samples; we call it the partial symmetrization technique. The method of Sherman (1994) uses only two samples, but includes products of m sign variables. The bounds obtained by Sherman's results are not as sharp as those of Arcones and Giné, because they are based on moment inequalities, yet in the case of increasing degree kernels they offer a viable alternative method of proof. We call it the complete symmetrization method.

Theorem 1 (Uniform weak law of large numbers—Partial Symmetrization) *Let the sequence $\{\mathcal{H}_m\}_{m=1}^{\infty}$ be the kernel class sequence of an infinite-degree U-process. Assume the following conditions:*

1. *The envelope H_m for $\mathcal{H}_m$ is P^m-integrable, and for each ε there exists a K_m such that*

$$\lim_{m\to\infty} P^m H_m\{H_m > K_m\} < \varepsilon.$$

2. *The class $\mathcal{H}_m$ is totally bounded under pseudometric $d_{1,j}$ for each j, m, and for every $\varepsilon \in (0,1]$,*

$$\frac{2^{m+1}K_m}{n^{1/2}}\mathbb{P}[\log N(\varepsilon^2/2^{m+3}K_m, d_{1,0}, \mathcal{H}_m)]^{1/2} + \frac{K_m}{n^{1/2}}\sum_{j=0}^{m-1}\mathbb{P}\binom{m}{j+1}\mathbb{P}[\log N(\varepsilon^2/8m\binom{m}{j+1}K_m, d_{1,j}, \mathcal{H}_m)]^{1/2} = o(1).$$

Then for every $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{\mathcal{H}_m}|U_{n,m}h - P^m h| > \varepsilon\right) \to 0.$$

Corollary 2 (For Euclidean classes) *Under the same conditions as in Theorem 1, assume there exist constants A and V, such that*

$$\max_{0\le j\le m-1} N(\varepsilon, d_{1,j}, \mathcal{H}_m) \le A\varepsilon^{-V},$$

and

$$\frac{m2^m K_m \log K_m}{n^{1/2}} = o(1),$$

then the uniform weak law of large numbers holds.

The next results require one of the following conditions on the metric entropy of the approximating classes $\mathcal{H}_{m(j)}$:

(A) The classes $\mathcal{H}_{m(j)}$ are Euclidean(A_j, V_j) for the envelopes $H_{m(j)} \le 2^j K_m$, $\max_j A_j \le A$, $\max_j V_j \le V_m^*$, and $K_m m^{1+k} V_m^{*(k+1)/2}/n^{1/2} = o(1)$, $k = 0, 1$.

(B) For a sequence of positive integers $\{r_j\}_{j=1}^m$, where

$$(r_j)^j/(r_{j-1})^{(j-1)} \le n^p \text{ for some } 0 \le p < 1,$$

and for $k = 0, 1$,

$$(2r_{k+1}-1)^{(k+1)/2}m^{k+1}n^{-1/2}\sum_{j=k+1}^m \mathbb{P}\left[\left(U_{2n,j}H_{m(j)}^2\right)^{1/2}\int_0^{1/4} N(x, d_{2,j}, \mathcal{H}_{m(j)})^{1/2r_j}dx\right] = o(1).$$

(C) For a sequence of positive integers $\{r_j\}_{j=1}^m$, $k = 0, 1$,

$$\sum_{j=k+1}^m \sqrt{j!}(16r_j - 8)^{j/2}n^{-j/2}\binom{m}{j} \mathbb{P}\left[\left(U_{2n,j}H_{m(j)}^2\right)^{1/2}\int_0^{1/4} N(x, d_{2,j}, \mathcal{H}_j)^{1/2r_j}dx\right] = o(n^{-k/2}).$$

Theorem 3 (Uniform weak law of large numbers—Complete symmetrization) *Assume the following conditions:*

1. *The envelope H_m for $\mathcal{H}_m$ is P^m-square integrable, and*
$$\lim_{m\to\infty} P^m H_m^2 < \infty.$$
2. *The classes $\mathcal{H}_{m(j)}$ are totally bounded under pseudometric $d_{2,j}$ for each j, m, and one of conditions (A), (B), or (C) holds for $k = 0$.*

Then for every $\varepsilon > 0$,
$$\mathbb{P}\left(\sup_{\mathcal{H}_m} |U_{n,m}h - P^m h| > \varepsilon\right) \to 0.$$

Theorem 4 (Rates of Convergence—Complete symmetrization) *Assume the following conditions:*

1. *The envelope H_m for $\mathcal{H}_m$ is P^m-square integrable, and*
$$\lim_{m\to\infty} P^m H_m^2 < \infty.$$
2. *The classes $\mathcal{H}_{m(j)}$ are totally bounded under pseudometric $d_{2,j}$ for each j, m, and one of conditions (A),(B), or (C) holds for $k = 1$.*

Then for every $\varepsilon > 0$,
$$\mathbb{P}\left(\sup_{\mathcal{H}_m} \sqrt{n}\left|U_{n,m}h - P^m h - mU_{n,1}h_{m(1)}\right| > \varepsilon\right) \to 0.$$

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
$$mh_{m(1)}(x) = \sum_{j=1}^{m} j\left[f_{j|1}(x) - P^j f_j\right],$$
and therefore
$$mU_{n,1}h_{m(1)} = (P_n - P)F_m,$$
where
$$F_m(x) \stackrel{\text{def}}{=} \sum_{j=1}^{m} jf_{j|1}(x).$$
It is assumed that the $\mathcal{F}_k$ have a common index set, say $\mathcal{T}$.

Theorem 5 *Suppose the following conditions hold:*

1. *For each $s, t \in \mathcal{T}$,*
$$\sigma(s,t) = \lim_{n\to\infty} \sum_{j=1}^{m}\sum_{k=1}^{m} jk\mathbb{P}\left[f_{j|1}(X_1;s) - Pf_{j|1}(;s)\right]\left[f_{k|1}(X_1;t) - Pf_{k|1}(;t)\right] | < \infty.$$

2. *The class $\{F_m\}$ is Euclidean(A, V) for the envelope $\bar{F}_m$, all m.*

3. $\limsup P\bar{F}_m(X)^2 < \infty$.

4. $\limsup P\bar{F}_m(X)^2\{\bar{F}_m > \varepsilon/n\} \to 0$ *for each* $\varepsilon > 0$.

5. *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 $\rho(s,t) = \lim_{n\to\infty}\rho_n(s,t)$ is well-defined, and for all deterministic sequences $\{s_n\}$ and $\{t_n\}$, if $\rho(s_n,t_n) \to 0$, then $\rho_n(s_n,t_n) \to 0$.

Then $n^{1/2}(P_n - P)F_m$ converges in distribution to a mean-zero Gaussian process which is uniformly ρ-continuous and has covariance given by $\sigma(s,t)$.

The proof of Theorem 1 appears in Section 5, proofs of Theorems 3 and 4 appear in Section 6, and a proof of Theorem 5 appears in the Appendix.

Comparison of conditions We have presented two sets of conditions for proving uniform weak laws of large numbers for infinite-degree U-processes. The conditions relate to the two techniques of proof discussed above. Both incur constraints on the rate at which m may grow, due to the decomposition of a sign-symmetrized version of $U_{n,m}h_m$ into 2^m components. When m is fixed, we recover results on finite-degree U-processes.

The requirements on the kernel classes appear as the second condition in each of the theorems. Two differences are obvious: For each m, the conditions stated in Theorems 1 applies to a single kernel class under m different metrics, while conditions A, B and C of Theorems 3 and 4 apply to m kernel classes under m metrics. Since the partial symmetrization results are derived using a one-step approximation and an exponential moment inequality, it places a constraint only on the square root of the logarithm of the covering numbers (although the precision of the cover needs to grow with m). In contrast, the complete symmetrization method follows from a combination of chaining and a weaker moment inequality, so the condition is an integral of a fractional power of the covering number. Condition (A) with $k = 0$ for Theorem 3 corresponds to the conditions of Corollary 2. Suppose the Euclidean parameter V_m^* is bounded by V^*, all j, m. (Our example in the next section meets these constraints.) Then condition (A) becomes

$$K_m m \log m/n^{1/2} \to 0.$$

Thus in examples in which it is easy to apply both results, the complete symmetrization technique allows m to grow more quickly than the partial symmetrization technique. In fact, the constraints of the partial symmetrization technique are so restrictive that we do not present a functional central limit theorem for it, because the $n^{1/2}$ normalization slows down m even further.

Theorem 1 can be modified to hold for more general kernel classes $\mathcal{G}$ with a small change in the metric that reflects the lack of symmetry in the kernel functions.

Related work 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 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 *et al* (1997) consider statistics based on resampling procedures. Among the procedures summarized in Bickel *et al* (1997), the m out of n bootstrap is an infinite-degree V-statistic and the $\binom{n}{m}$ bootstrap is an infinite-degree U-statistic. Grübel and Pitts (1993) and Harel *et al* (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.

3 Example

Recall from the Introduction that in our example the $\{X_i\}_{i=1}^n$ are i.i.d. random variables with probability measure P. Frees proposed the following U-statistic as an estimator for the renewal measure,

$$\begin{aligned}\hat{N}_m(t) &= 1 + \sum_{k=1}^{m} U_{n,k} f_{k,t} \\ &= 1 + U_{n,m} g_{m,t},\end{aligned}$$

where $f_{k,t}$ is the subkernel and $g_{m,t}$ is the grand kernel:

$$f_{k,t}(x_1, \ldots, x_k) = \{x_1 + \cdots + x_k \le t\},$$

$$g_{m,t}(x_1, \ldots, x_m) = \sum_{k=1}^{m} f_{k,t}(x_1, \ldots, x_k).$$

While the renewal estimator $\hat{N}_m$ is biased for each m, its expectation converges to the renewal function N, as $m(n) \to \infty$. This estimator is seen to be a U-statistic of degree m, or alternatively a sum of U-statistics with kernels of degree $k = 1, \ldots, m$. While each of the subkernels f_k is permutation-symmetric, the grand kernel g_m is not symmetric. A symmetric version of the renewal estimator is:

$$1 + U_{n,m} h_{m,t},$$

where

$$h_{m,t}(x_1, \ldots, x_m) = \frac{1}{m!} \sum_{\pi \in (m)_m} g_{m,t}(x_{\pi_1}, \ldots, x_{\pi_k}).$$

Before stating the asymptotic results for the renewal function estimator, we consider the entropy properties of the index classes. First, define the graph $\Gamma(f)$ for a real-valued function f on $\mathcal{S}^k$:

$$\Gamma(f) = \{(\boldsymbol{x}, y) \in S^k \times \mathbb{R} : 0 \le y \le f(\boldsymbol{x}) \text{ or } f(\boldsymbol{x}) \le y \le 0\}.$$

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 $\mathcal{F}_k$ is Euclidean$(A, 2)$ for the envelope 1, that $\mathcal{G}_m$ is Euclidean$(A, 2)$ for the envelope m, 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 $2^i m$ for some constant A.

To prove this, we use Lemma II.25 of Pollard (1984), which states that if the graphs of functions $f \in \mathcal{F}$ can not shatter any set of k points (i.e. pick out all 2^k subsets via intersection) then $\mathcal{F}$ is Euclidean for its natural envelope, and $V = 2(k-1)$. Consider the collection of graphs $\{\Gamma(f) : f \in \mathcal{F}_k\}$. We show that these graphs can shatter no set of two points. Let $z, z' \in S^k \times \mathbb{R}$. Proceed by contradiction, and suppose the graphs can shatter these two points: that is, there exists $t_1 < t_2 < t_3 < t_4 \in \mathbb{R}$ and their corresponding graphs $\Gamma(f_1)$, $\Gamma(f_2)$, $\Gamma(f_3)$, $\Gamma(f_4)$ such that $\Gamma(f_{i_1})$ picks out the empty set, $\Gamma(f_{i_2})$ picks out z', $\Gamma(f_{i_3})$ picks out z, and $\Gamma(f_{i_4})$ picks out both z and z'. In more compact notation, we have $z \in \Gamma(f_{i_1})^c\Gamma(f_{i_2})^c\Gamma(f_{i_3})\Gamma(f_{i_4})$ and $z' \in \Gamma(f_{i_1})^c\Gamma(f_{i_2})\Gamma(f_{i_3})^c\Gamma(f_{i_4})$. Since $t_1 < t_2 < t_3 < t_4$, the four graphs are subsets of each other: $\Gamma(f_1) \subseteq \Gamma(f_2) \subseteq \Gamma(f_3) \subseteq \Gamma(f_4)$. Consider all intersections of the form $\Gamma(f_{i_1})^c\Gamma(f_{i_2})^c\Gamma(f_{i_3})\Gamma(f_{i_4})$, and note that the subset property implies that $\Gamma(f_1)^c\Gamma(f_2)^c\Gamma(f_3)\Gamma(f_4)$ is uniquely nonempty. Thus these graphs cannot shatter z and z'. It then follows that the class $\mathcal{F}_k$ is Euclidean$(A, 2)$ for the envelope 1.

The entropy of $\mathcal{F}_k$ was determined by the ordering property alone. Since this same property carries over to the graphs of $g \in \mathcal{G}_m$, and $h \in \mathcal{H}_m$, they too are Euclidean$(A, 2)$ for the constant envelope m. This property carries over to conditional expectations as well, and so $\mathcal{H}_{m|k}$ is Euclidean$(A, 2)$ for the envelope m. Additionally, for the projected kernel,

$$\begin{aligned} h_{m(j)}(x_1, \ldots, x_j) &= \left[\sum_{k=0}^{j}\{j-k \text{ is even}\} \sum_{\pi \in (j,k)} h_{m|k}(x_{\pi_1}, \ldots, x_{\pi_k})\right] \\ &\quad - \left[\sum_{k=0}^{j}\{j-k \text{ is odd}\} \sum_{\pi \in (j,k)} h_{m|k}(x_{\pi_1}, \ldots, x_{\pi_k})\right], \end{aligned}$$

each of the sums in square brackets is Euclidean for the envelope $2^{(j-1)}m$. This means that $\mathcal{H}_{m(j)}$ is Euclidean$(A, 4)$ for the envelope $2^j m$ (Corollary 17, Nolan and Pollard, 1987). Results such as these can be found in Dudley (1985, p 496 and Example 5.4, p 506).

Now we are ready to state our results. The proofs are found in the Appendix.

Lemma 6 *Let P be a probability measure with positive mean, and let τ be a finite constant. Then for any $m(n) \leq n$ such that $m(n) \to \infty$,*

$$\sup_{t \in [0,\tau]} \left|\hat{N}_m(t) - N(t)\right| \xrightarrow{\mathbb{P}} 0.$$

Lemma 7 *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 $m(n) \geq n^{1/(6+2\eta)}$, 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*

$$\lim_{n \to \infty} \sum_{j=1}^{m}\sum_{k=1}^{m} jk[\mathbb{P}P^{[j-1]}(s-x)P^{[k-1]}(t-x) - P^{[j]}(s)P^{[k]}(t)].$$

Comments In contrast to our approach, Grübel and Pitts (1993) study the infinite-degree V-process

$$V_{n,n}f_n = \sum_{\pi \in \{1,\ldots,n\}^n} f_{n,t}(X_{\pi_1}, \ldots, X_{\pi_n}).$$

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 *et al* (1995) study the same estimator, restricted to non-negative random variables. Schneider *et al* (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) advise that Frees' estimator may be more suitable when the sample size is small, one is interested in relatively small values of t, and the renewal times are positive.

4 Preliminaries

We collect here some inequalities that are required for the proofs of Theorems 1, 3, and 4. The moment inequality is found in Bonami (1970), the exponential inequality dates to Pisier (1983) and Giné and Zinn (1984).

Lemma 8 (Convex maximal inequality) *Let Ψ be a convex, strictly increasing function on $[0,\infty)$. Let p be a positive integer. Suppose $Z_1,\ldots,Z_N$ are random variables and Δ is a positive real number such that $\mathbb{P}\Psi\left(|Z_l|^p/\Delta^p\right)\leq 1$ for each l. Then*

$$\left[\mathbb{P}\max_{l\leq N}|Z_l|^p\right]^{1/p}\leq\Delta\left[\Psi^{-1}(N)\right]^{1/p}. \tag{5}$$

Lemma 9 (Exponential inequality) *Let $\{u_l(\cdot)\}_{l=1}^N$ be a sequence of mappings taking one argument, and let $\{v_l(\cdot,\cdot)\}_{l=1}^N$ be a sequence of mappings taking two arguments. Then for some finite constant C not depending on n, the following maximal inequalities hold:*

$$\mathbb{P}\max_{l\leq N}\left|\sum_{i=1}^n\sigma_i u_l(i)\right|\leq C(\log N)^{1/2}\max_{l\leq N}\left[\sum_{i=1}^n u_l(i)^2\right]^{1/2}.$$

Lemma 10 (Moment inequality) *Let $u(\pi_1,\ldots,\pi_k)$ be a mapping that is permutation-symmetric in its k indices. For every integer $s\geq 1$, we have*

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

Lemma 11 (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:*

1. *If $d(s,t)=0$, then $Z(s)=Z(t)$ almost surely.*
2. *If $d(s,t)>0$, then $\mathbb{P}\Psi(|Z(s)-Z(t)|^p/d(s,t)^p)\leq 1$.*
3. *There exists a point t_0 in T for which $\delta\overset{def}{=}\sup_T d(t,t_0)<\infty$.*
4. *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.

5 Partial symmetrization technique

In this section we establish the proof of Theorem 1 via the partial sign symmetrization technique. We first establish the following symmetrization lemma, which uses independent sign variables, $\sigma_1, \ldots, \sigma_n$, where $\sigma_i = \pm 1$ with equal probability.

Lemma 12 (Partial Symmetrization Inequality)

$$\begin{aligned}\mathbb{P}\sup_{\mathcal{H}} |U_{n,m}h - P^m h| &\leq 2^{m+1}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}\sigma_{\pi_1}h(\pi,0)\right| \\ &+ 2\sum_{j=1}^{m-1}\binom{m}{j}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}\sigma_{\pi_1}h(\pi,j)\right|.\end{aligned}$$

Proof. We begin by introducing three more samples, $Y_1, \ldots, Y_n$, $\tilde{Y}_1, \ldots, \tilde{Y}_n$, and $Y'_1, \ldots, Y'_n$, where the Ys are independent samples from P that are also independent of the three X samples. Next we introduce notation for conditional expectations. Let $\mathbb{P}_{\mathcal{X}}$ denote expectation conditional on $X_1, \ldots, X_n$. Let $\mathbb{P}_{\mathcal{Z}}$ denote conditioning on the unordered pairs $(X_1, X'_n), \ldots, (X_n, X'_n)$, which is equivalent to conditioning on $Z_1, \ldots, Z_n$ where $Z_i = (X_i, X'_i)$ with probability $1/2$ and $Z_i = (X'_i, X_i)$ with probability $1/2$.

We now present some inequalities that will lead to the establishment of the lemma. They are in the same spirit as the work of de la Peña (1992), but result in factors such as 2^m rather than de la Peña's

$$2^m(m^m-1)((m-1)^{(m-1)}-1)\times\cdots\times(2^2-1),$$

which is too large for our purposes.

$$\begin{aligned}\mathbb{P}_{\mathcal{Z}}h(X_1,\ldots,X_m) &= \mathbb{P}_{\mathcal{Z}}h(X'_1,\ldots,X'_m) \\ &= \mathbb{P}_{\mathcal{Z}}h(X_1,X'_2,\ldots,X'_m) \\ &= \frac{1}{2^m}\sum_{b=0}^{2^m-1}h(b_1X_1+(1-b_1)X'_1,\ldots,b_mX_m+(1-b_m)X'_m), \qquad (6)\end{aligned}$$

where the summation is over integers $b = 0, 1, \ldots, 2^m - 1$ with $(b_1, \ldots, b_m)$ being the binary expansion of b. The last equality results from noting the sum is measurable with respect to $\mathcal{Z}$. Recall the definition of $h(\pi, j)$ from section 2 in (4), and expand it slightly as follows,

$$h_m(\pi,j,X',X,\tilde{X}) = h(X'_{\pi_1}, X_{\pi_2},\ldots,X_{\pi_{j+1}},\tilde{X}_{\pi_{j+2}},\ldots,\tilde{X}_{\pi_m}).$$

Similarly define $h(\pi,j,Y',Y,\tilde{Y})$, $h(\pi,j,X',Y,\tilde{Y})$, and $h(\pi,j,Y',X,\tilde{X})$. The following two sums have the same distribution.

$$\frac{1}{(n)_m}\sum_{\pi\in(n)_m}h(\pi,j,X',X,\tilde{X})-h(\pi,j,Y',Y,\tilde{Y}) \sim \frac{1}{(n)_m}\sum_{\pi\in(n)_m}\sigma_{\pi_1}[h(\pi,j,Y',X,\tilde{X})-h(\pi,j,X',Y,\tilde{Y})]. \quad (7)$$

We will use these properties to establish the lemma. First add, for $b = 1, \ldots, 2^m - 2$, each of the following terms,

$$\mathbb{P}_{\mathcal{X}}h(\pi,|b|,X',X,X') - \mathbb{P}_{\mathcal{X}}h(\pi,|b|,X',X,\tilde{X}) = 0,$$

and apply the triangle inequality.

$$
\begin{aligned}
&\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m} h(X_{\pi_1},\ldots X_{\pi_m}) - P^m h\right| \\
&\leq \mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m} h(X_{\pi_1},\ldots,X_{\pi_m}) - P^m h + \sum_{b=1}^{2^m-2}\mathbb{P}_{\mathcal{X}}[h(\pi,|b|,X',X,X') - P^m h]\right| \\
&\quad + \sum_{b=1}^{2^m-2}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}\mathbb{P}_{\mathcal{X}}[h(\pi,|b|,X',X,\tilde{X}) - P^m h]\right|.
\end{aligned}
$$

Next bound this upper bound by applying (6) to the first term. Then apply Jensen's inequality conditionally on the three X samples, averaging over the three Y samples.

$$
\begin{aligned}
&\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m} h(X_{\pi_1},\ldots X_{\pi_m}) - P^m h\right| \\
&\leq 2^m\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}[h(\pi,1,X',X,X') - P^m h]\right| \\
&\quad + \sum_{j=1}^{m-1}\binom{m}{j}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}\mathbb{P}_{\mathcal{X}}[h(\pi,j,X',X,\tilde{X}) - P^m h]\right| \\
&\leq 2^m\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}[h(\pi,1,X',X,X') - h(\pi,1,Y',Y,Y')]\right| \\
&\quad + \sum_{j=1}^{m-1}\binom{m}{j}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}[h(\pi,j,X',X,\tilde{X}) - h(\pi,j,Y',Y,\tilde{Y})]\right|
\end{aligned}
$$

Finally apply (7) to obtain the bound,

$$
2^{m+1}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}\sigma_{\pi_1}h(\pi,0)\right| + 2\sum_{j=1}^{m-1}\binom{m}{j}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}\sigma_{\pi_1}h(\pi,j)\right|.
$$

The Lemma has now been established. $\square$

Proof of Theorem 1. From Markov's inequality and Lemma 12 we see that

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\mathcal{H}}|U_{n,m}h - P^m h| > \varepsilon\right) \\
&\leq \frac{2^{m+1}}{\varepsilon}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}\sigma_{\pi_1}h(\pi,0)\right| \\
&+ \frac{2}{\varepsilon}\sum_{j=1}^{m-1}\binom{m}{j}\mathbb{P}\sup_{\mathcal{H}}\left|\frac{1}{(n)_m}\sum_{\pi\in(n)_m}\sigma_{\pi_1}h(\pi,j)\right|.
\end{aligned}
$$

For each j, construct an approximating class $\mathcal{H}'(j)$ to $\mathcal{H}_m$ using the pseudometric $d_{1,j}$ defined in Section 2. Then for every h in $\mathcal{H}$ there is an $h' \in \mathcal{H}'(j)$ such that $d_{1,j}(h,h') \leq \varepsilon_j$, where ε_j will be determined later. This implies

$$\begin{aligned}
&\sup_{\mathcal{H}} \left| \frac{1}{(n)_m} \sum_{\pi \in (n)_m} \sigma_{\pi_1} [h(\pi, j) - h'(\pi, j)] \right| \\
&\leq \sup_{\mathcal{H}} \frac{1}{(n)_m} \sum_{i=1}^{n} \left| \sum_{\pi \in (n)_m : \pi_1 = i} [h(\pi, j) - h'(\pi, j)] \right| \\
&\leq \varepsilon_j \frac{1}{(n)_m} \sum_{\pi \in (n)_m} H_m(\pi, j).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{P} \sup_{\mathcal{H}} \left| \frac{1}{(n)_m} \sum_{\pi \in (n)_m} \sigma_{\pi_1} h(\pi, j) \right| \\
&\leq \varepsilon_j P^m H_m + \mathbb{P} \max_{\mathcal{H}'(j)} \left| \sum_{i=1}^{n} \sigma_i \left| \frac{1}{(n)_m} \sum_{\pi \in (n)_m : \pi_1 = i} h'(\pi, j) \right| \right|.
\end{aligned}$$

Apply the exponential inequality (Lemma 9) to the second piece on the right-hand side (conditioned on the triple X sample) with

$$u_l(i) = \left| \frac{1}{(n)_m} \sum_{\pi \in (n)_m : \pi_1 = i} h'(\pi, j) \right| \leq K_m / n.$$

This gives the bound

$$\frac{K_m}{n^{1/2}} \mathbb{P} C [\log N(\varepsilon_j, d_{1,j}, \mathcal{H}_m)]^{1/2}.$$

Take ε_0 to be $\varepsilon^2/8K_m 2^m$ and ε_j to be $\varepsilon^2/8K_m m\binom{m}{j+1}$, $j = 1, \ldots m-1$, to find the desired result. In the case of an unbounded envelope, truncate each h to $\{H_m < K_m\}$. Then the envelope $H_m\{H_m < K_m\}$ is bounded by K_m. The entropy of the truncated class is smaller than the entropy of the original class $\mathcal{H}_m$, and K_m can be chosen such that the error incurred by the truncation is at most ε. $\square$

6 Complete symmetrization technique

For the complete symmetrization, 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. The difference between the complete and partial sign-symmetrization is that now 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. The maximal inequality is applied to each component of the Hoeffding decomposition.

Proof of Theorems 3 and 4 To begin, consider a class of kernels $\mathcal{H}$, degenerate and of degree j. For an integer $b = 0, 1, \ldots, 2^j - 1$, let $(b_1, \ldots, b_j)$ be its binary expansion, and for each $h \in \mathcal{H}$, define for $\pi \in (n)_j$,

$$h^{\circ}(\boldsymbol{X}_\pi, \boldsymbol{X}'_\pi) \stackrel{\text{def}}{=} \sum_{b=0}^{2^j-1} (-1)^{\sum b_i} h(b_1 X_{\pi_1} + (1-b_1) X'_{\pi_1}, \ldots, b_j X_{\pi_j} + (1-b_j) X'_{\pi_j}).$$

Since h is degenerate, we have

$$\mathbb{P}\left[h^{\circ}(\boldsymbol{X}_{\pi}, \boldsymbol{X}'_{\pi}) \middle| X_1, \ldots, X_n\right] = h(X_{\pi_1}, \ldots, X_{\pi_j}).$$

Use this fact and Jensen's inequality (applied conditionally) to obtain the upper bound,

$$\begin{aligned} &\mathbb{P} \sup_{\mathcal{H}} |U_{n,j} h|^p \\ \leq \quad &\mathbb{P} \sup_{\mathcal{H}} \left| \frac{1}{(n)_j} \sum_{\pi \in (n)_j} h^{\circ}(\boldsymbol{X}_{\pi}, \boldsymbol{X}'_{\pi}) \right|^p . \end{aligned}$$

It also follows that

$$\mathbb{P} \sup_{\mathcal{H}} |U_{n,j} h|^p \leq \mathbb{P} \sup_{\mathcal{H}} \left| U^{\circ}_{n,j} h \right|^p ,$$

where

$$U^{\circ}_{n,j} h \stackrel{\text{def}}{=} \frac{1}{(n)_j} \sum_{\pi \in (n)_j} \sigma_{\pi_1} \cdots \sigma_{\pi_j} h^{\circ}(\boldsymbol{X}_{\pi}, \boldsymbol{X}'_{\pi}).$$

This inequality allows us to consider the completely sign-symmetrized process $U^{\circ}_{n,j}$ in place of the original degenerate process $U_{n,j}$.

We now extend Sherman's maximal inequality to infinite-degree U-processes. The maximal inequality is then applied to the components of the Hoeffding decomposition. Recall that $U_{2n,m}$ is the U-statistic applied to the double sample relabeled as $W_i = X_i$, and $W_{n+i} = X'_i$, $i = 1, \ldots, n$.

Theorem 13 (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*

$$\begin{aligned} \tau_{n,j} &= \left(U_{2n,j} H^2\right)^{1/2} \\ \delta_{n,j} &= \frac{\sup_{\mathcal{H}} \left(U_{2n,j} h^2\right)^{1/2}}{4\left(U_{2n,j} H^2\right)^{1/2}} \\ \Gamma_j &= \left[64 j! (16pr - 8)^j\right]^{p/2} . \end{aligned}$$

Then

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

Proof. We will apply the chaining inequality (Lemma 11) to a normalized version of the sign-symmetrized U-process $U^{\circ}_{n,j}$. In Lemma 11, let the stochastic process $Z(t)$ be $n^{j/2} U^{\circ}_{n,j} h / \tau_{n,j}$; let the pseudometric be $d_{2,j}(h, h')$ defined in Section 2; and, let the convex function $\Psi(x)$ be x^{2r}/γ. The factor γ will be chosen in order to satisfy the second condition of the chaining lemma, which we establish now. Condition on the double sample $\mathcal{W}$ to find the bound

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

Inequality (8) follows from

$$n^{j/2}(2n)_j^{1/2}(n)_j^{-1} < 2^j \text{ for } j < n/3.$$

Next apply Bonami's inequality (Lemma 10) to the upper bound in (8) with

$$u(\pi_1, \ldots, \pi_j) = \frac{(h^\circ - h'^\circ)(X_{\pi_1}, \ldots, X_{\pi_j}, X'_{\pi_1}, \ldots, X'_{\pi_j})}{\left[\sum_{i \in (2n)_j} (h - h')(W_{i_1}, \ldots, W_{i_j})^2\right]^{1/2}}.$$

This yields the upper bound,

$$4^{jpr}\left[j!^{pr}(2pr-1)^{jpr}\right]\mathbb{P}_{\mathcal{W}}\left[\frac{\sum_{\pi \in (n)_j}(h^\circ - h'^\circ)(\boldsymbol{X}_\pi, \boldsymbol{X}'_\pi)^2}{\sum_{i \in (2n)_j}(h-h')(W_{i_1}, \ldots, W_{i_j})^2}\right]^{pr}.$$

Finally, apply the following inequality,

$$\sum_{\pi \in (n)_j}\left[h^\circ(\boldsymbol{X}_\pi, \boldsymbol{X}'_\pi)\right]^2 \le 2^j \sum_{i \in (2n)_j} h(W_{i_1}, \ldots, W_{i_j})^2, \tag{9}$$

to obtain the bound,

$$\mathbb{P}_{\mathcal{W}}\left[\frac{n^{j/2}U^\circ_{n,j}(h-h')}{\left[U_{2n,j}(h-h')^2\right]^{1/2}}\right]^{2pr} \le [j!(16pr-8)^j]^{pr}.$$

(Inequality (9) is established in the Appendix.) Letting $\gamma = \left[j!(16pr-8)^j\right]^{pr}$, 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, by (9) we have

$$\left|U^\circ_{n,j}(h-h')\right| \le \left[\frac{2^j(2n)_j}{(n)_j}U_{2n,j}(h-h')^2\right]^{1/2} < 2^j d_{2,j}(h,h').$$

This inequality also gives us the first condition.

With $\Psi^{-1}(y) = (\gamma y)^{1/2r}$, the chaining lemma gives

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

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

The maximal inequality is the centerpiece for 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 $\mathcal{H}_m$ a sequence of kernel classes of an infinite-degree U-process, with associated envelopes H_m where $P^m H_m < \infty$. For $j = 1, \ldots, m$, 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 $|h_{m|k}| \le H_{m|k}$ (a convenient abuse of the conditioning notation), then for every $h \in \mathcal{H}$,

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

Then, by Theorem 13, bound the jth Hoeffding projection:

$$\mathbb{P} \sup_{\mathcal{H}_{(m(j)}} \left| \binom{m}{j} U_{n,j} h_{m(j)} \right| \leq n^{-j/2} \binom{m}{j} \Gamma_j \mathbb{P} \left[\left(U_{2n,j} H^2_{m(j)} \right)^{1/2} J(j, r_j) \right],$$

where now $\Gamma_j = \left[64j!(16r_j - 8)^j\right]^{1/2}$, and $J(j,r)$ is the covering integral for the class $\mathcal{H}_{m(j)}$ with envelope $H_{m(j)}$ under metric $d_{2,j}$:

$$J(j,r) = \int_0^{1/4} N(x, d_{2,j}, \mathcal{H}_{m(j)})^{1/2r} dx.$$

The upper limit of $1/4$ for the covering integral is valid, since $\delta_{n,j} \leq 1/4$. Sum over j and Theorems 3 and 4 follow under condition (C) because, for $k = 0, 1$,

$$\mathbb{P} \sup_{\mathcal{H}} \left| \sum_{j=k+1}^{m} \binom{m}{j} U_{n,j} h_{m(j)} \right| = o(n^{-k/2}).$$

As for the proof under conditions (A) and (B), note that (A) $\rightarrow$ (C) and (B) $\rightarrow$ (C). (A proof appears in the Appendix). □

7 Appendix

Proof of Theorem 5 The proof simply invokes a version of Pollard's (1990) Theorem 10.6 with

$$f_{ni} = n^{-1/2} F_m(X_i; t)$$
$$F_{ni} = n^{-1/2} \bar{F}_m(X_i)$$
$$X_n(t) = n^{1/2}(P_n - P) F_m(; t)$$

The first condition implies the first two conditions in Pollard (1990) Theorem 10.6, which gives the finite dimensional convergence. The second condition is actually slightly stronger than the corresponding condition in Pollard (1990), but for simplicity, it is stated in a form that suits our purposes. □

Proof of Lemma 6 The condition that $m(n) \rightarrow \infty$ implies,

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

Let $m^* = \log n/4$. For any $m(n) \geq m^*$, note that

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

Therefore, we need only consider $m(n) \leq \log n/4$. Finally, apply the Corollary to Theorem 1 to the IDUP,

$$\hat{N}_m(t) - \mathbb{P}\hat{N}_m(t) = U_{n,m} g_{m,t} - P^m g_{m,t}.$$

The IDUP is indexed by the grand kernel $\mathcal{G}_m$, which has been shown to be Euclidean$(A, 1)$ for the envelope m. The conditions of the theorem are easily met. □

Proof of Lemma 7 As in the proof of Lemma 6, we replace N by $\mathbb{P}\hat{N}_m$. To do this replacement, we use the following result (Gut, 1974, Theorem 2.1): for $r > 2$, $\mathbb{P}\left|X^-\right|^r < \infty$ implies

$$\sum_{k\geq 1} k^{r-2} P^{[k]}(t) < \infty. \tag{10}$$

Therefore, for $m \geq m^* = n^{1/(2r-4)}$,

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

and,

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

Now we can work with the IDUP,

$$\hat{N}_{m^*}(t) - \mathbb{P}\hat{N}_{m^*}(t) = U_{n,m^*} g_{m^*,t} - P^{m^*} g_{m^*,t}.$$

We wish to show that Theorem 4 holds, i.e. that for every $\varepsilon > 0$,

$$\mathbb{P}\left(n^{1/2} \sup_{t\in T} \left| U_{n,m^*}h - P^{m^*}h - (P_n - P)F_{m^*,t} \right| > \varepsilon \right) \to 0,$$

where

$$F_{m^*,t}(x) = \sum_{j=1}^{m^*} j P^{[j-1]}(-\infty, t-x].$$

This follows from condition (A) provided $m^* = o(n^{1/6})$. When $r > 5$, this constraint on m^* is satisfied. (See the the condition in the Lemma on the rth moment of $|X^-|$.)

Finally, for the functional central limit theorem, apply Theorem 5 to the first-order projection $n^{1/2}(P_n - P)F_m(t)$. Use (10) and the following bound to verify the first and fourth conditions of the Theorem.

$$\begin{aligned} &\sum_{j=1}^{m}\sum_{k=1}^{m} jk \left[\mathbb{P} P^{[j-1]}(-\infty, s-X_1] P^{[k-1]}(-\infty, t-X_1] - P^{[j]}(-\infty, s] P^{[k]}(-\infty, t] \right] \\ \leq \quad & \sum_{j=1}^{m}\sum_{k=1}^{m} jk P^{[j]}(s) \wedge P^{[k]}(t). \end{aligned}$$

In addition let, $F_m(x;\tau)$, be the envelope for $\{F_m\}$. Since $\mathcal{F}_k$ is Euclidean$(A,1)$ for envelope 1, the collection $\{F_m\}$ is Euclidean$(A,1)$ and $P^m F_m(\cdot;\tau)^2 < \infty$. All the conditions of Theorem 5 are met. The result now follows. □

Proof of (9) Expand the summands on the left side for each $\pi \in (n)_k$ and apply the Cauchy-Schwarz inequality:

$$\begin{aligned} \left[H_t^\circ(\boldsymbol{X}_\pi, \boldsymbol{X}'_\pi)\right]^2 &= \left[\sum_{b=0}^{2^j-1} (-1)^{\sum b_i} H_t(W_{\pi_1+b_1 n}, \ldots, W_{\pi_j+b_j n}) \right]^2 \\ &\leq 2^j \sum_{b=0}^{2^j-1} H_t(W_{\pi_1+b_1 n}, \ldots, W_{\pi_j+b_j n})^2. \end{aligned}$$

Sum over $\pi \in (n)_j$. In the double sum $\sum_{\pi\in(n)_j}\sum_{b=0}^{2^j-1}$, the aggregate index set of summation is

$$\left\{(\pi_1 + b_1 n, \ldots, \pi_j + b_j n) : \pi \in (n)_j,\, b = 0, \ldots, 2^j - 1\right\}.$$

Since each π_i is between 1 and n, adding an extra n to the subscripts does not introduce any repeated j-tuples. While each index now ranges from 1 to $2n$, the j-tuple of indices for each term is constrained so that no two indices are equal nor exactly n apart:

$$\left\{(i_1, \ldots, i_j) : \boldsymbol{i} \in \{1, \ldots, 2n\}^j,\, a \neq b \Rightarrow |i_a - i_b| \notin \{0, n\}\right\},$$

which is the same as

$$\{(i_1, \ldots, i_j) : \boldsymbol{i} \in (2n)_j,\, |i_a - i_b| \neq n \text{ for all } 1 \leq a, b \leq j\}.$$

Enlarge the index set of summation to $(2n)_j$ to conclude the proof. □

Proof of (B) → (C) Condition C bounds

$$\sum_{j=k+1}^{m} n^{-j/2}\binom{m}{j} j!^{1/2} 8^{j/2}(2r_j - 1)^{j/2}\mathbb{P}\left[\left(U_{2n,j}H^2_{m(j)}\right)^{1/2} J_j\right]. \tag{11}$$

Let $\phi_n(j) = n^{-j/2}\binom{m}{j} j!^{1/2} 8^{j/2}(2r_j - 1)^{j/2}$, and consider the ratio of successive terms:

$$\frac{\phi_n(j)}{\phi_n(j-1)} = \frac{8^{1/2}(m-j+1)(2r_j - 1)^{j/2}}{(jn)^{1/2}(2r_{j-1} - 1)^{(j-1)/2}},\; j = k+2, \ldots, m,$$

which is ≤ 1 for $j = k+2, \ldots, m$ if for each j,

$$m \leq \frac{(jn)^{1/2}(2r_{j-1} - 1)^{(j-1)/2}}{8^{1/2}(2r_j - 1)^{j/2}} + j - 1. \tag{12}$$

Then under the condition on the r_j, the bound (12) implies

$$m \leq \frac{n^{(1-p)/2}}{8^{1/2}}.$$

Thus for $j \geq k+2$, $\phi_n(j) \leq \phi_n(k+1)$, and we bound (11) by

$$Cn^{-(k+1-p)/2}(m)_{k+1} \sum_{j=k+1}^{m} \mathbb{P}\left[\left(U_{2n,j}H^2_{m(j)}\right)^{1/2} J_j\right].$$

The assumptions of condition B imply that condition C is met. The proof that condition A implies C follows similarly. We take $r_j = V_m^* + 1$, and note that if $|H_m| \leq K_m$ then $(P^j H^2_{m(j)})^{1/2} \leq 2^j K_m$. The factor of 2^j is easily absorbed into $\phi(j)$. □

References

[1] M. A. Arcones and E. Giné. Limit theorems for U-processes. *The Annals of Probability*, 21(3):1494–1542, 1993.

[2] P. J. Bickel, F. Götze, and W. R. van Zwet. Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7(1):1–31, 1997.

[3] G. Blom. Some properties of incomplete U-statistics. *Biometrika*, 63(3):573–580, 1976.

[4] E. Bolthausen and F. Götze. The rate of convergence for multivariate sampling statistics. *The Annals of Statistics*, 21(4):1692–1710, 1993.

[5] A. Bonami. tude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann Inst Fourier*, 20(2):335–402, 1970.

[6] V. H. de la Peña. Decoupling and Khintchine's inequalities for U-statistics. *The Annals of Probability*, 20(4):1877–1892, 1992.

[7] R. M. Dudley. A Course on Empirical Processes. In *cole d't de probabilits de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math*, pages 1–142. Springer, New York, 1984.

[8] R. M. Dudley. The structure of some Vapnik-Červonenkis classes. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume II, pages 495–508, Belmont, Calif, 1985. Wadsworth.

[9] E. B. Dynkin and A. Mandelbaum. Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *The Annals of Statistics*, 11(3):739–745, 1983.

[10] E. W. Frees. Nonparametric renewal function estimation. *The Annals of Statistics*, 14(4):1366–1378, 1986.

[11] E. W. Frees. Infinite order U-statistics. *Scandinavian Journal of Statistics*, 16:29–45, 1989.

[12] E. Giné and J. Zinn. Some limit theorems for empirical processes. with discussion. *The Annals of Probability*, 12(4):929–998, 1984.

[13] R. Grübel and S. M. Pitts. Nonparametric estimation in renewal theory I: the empirical renewal function. *The Annals of Statistics*, 21(3):1431–1451, 1993.

[14] A. Gut. On the moments and limit distributions of some first passage times. *The Annals of Probability*, 2(2):277–308, 1974.

[15] M. Harel, C. A. O'Cinneide, and H. Schneider. Asymptotics of the sample renewal function. *Journal of Mathematical Analysis and Applications*, 189:240–255, 1995.

[16] C.M. Heilig. *An Empirical Process Approach to U-processes of Increasing Degree.* Ph.D. dissertation, University of California, Berkeley.

[17] C.M. Heilig, and D. Nolan. Applications of the infinite-degree U-process. Manuscript in preparation.

[18] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19:293–325, 1948.

[19] A. Kohatsu-Higa. Weak convergence of infinite order U-processes. *Statistics & Probability Letters*, 12:145–150, 1991.

[20] A. J. Lee. *U-statistics: Theory and Practice.* Marcel Dekker, New York, 1990.

[21] D. Nolan and D. Pollard. U-processes: rates of convergence. *The Annals of Statistics*, 15(2):780–799, 1987.

[22] D. Nolan and D. Pollard. Functional limit theorems for U-processes. *The Annals of Probability*, 16(3):1291–1298, 1988.

[23] G. Pisier. Some applications of the metric entropy condition to harmonic analysis. In *Banach Spaces, Harmonic Analysis, and Probability Theory: Proceedings of the Special Year in Analysis*, volume 995 of *Lecture Notes in Math*, pages 123–154. Springer, New York, 1983.

[24] D. N. Politis and J. P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4):2031–2050, 1994.

[25] D. Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, New York, 1984.

[26] D. Pollard. *Empirical Processes: Theory and Applications*, volume 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, 1990.

[27] H. Schneider, B.-S. Lin, and C. O'Cinneide. Comparison of nonparametric estimators for the renewal function. *Journal of the Royal Statistical Society, Series C*, 39(1):55–61, 1990.

[28] R. J. Serfling. *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York, 1980.

[29] R. P. Sherman. Maximal inequalities for degenerate U-processes with applications to optimization estimators. *The Annals of Statistics*, 22(1):439–459, 1994.

[30] G. S. Shieh. Infinite order V-statistics. *Statistics & Probability Letters*, 20:75–80, 1994.