On weak approximation of $U$-statistics

Masoud M. Nasari

School of Mathematics and Statistics, Carleton University, Canada

e-mail: mmnasari@connect.carleton.ca

Abstract

This paper investigates weak convergence of $U$-statistics via approximation in probability. The classical condition that the second moment of the kernel of the underlying $U$-statistic exists is relaxed to having $\frac{4}{3}$ moments only (modulo a logarithmic term). Furthermore, the conditional expectation of the kernel is only assumed to be in the domain of attraction of the normal law (instead of the classical two-moment condition).

1 Introduction

Employing truncation arguments and the concept of weak convergence of self-normalized and studentized partial sums, which were inspired by the works of Csörgő, Szyszkowicz and Wang in [5], [4], [2] and [3], we derive weak convergence results via approximations in probability for pseudo-self-normalized $U$-statistics and $U$-statistic type processes. Our results require only that (i) the expected value of the product of the kernel of the underlying $U$-statistic to the exponent $\frac{4}{3}$ and its logarithm exists (instead of having 2 moments of the kernel), and that (ii) the conditional expected value of the kernel on each observation is in the domain of attraction of the normal law (instead of having 2 moments). Similarly relaxed moment conditions were first used by Csörgő, Szyszkowicz and Wang [5] for $U$-statistics type processes for changepoint problems in terms of kernels of order 2 (cf. Remark 5). Our results in this exposition extend their work to approximating $U$-statistics with higher order kernels. The thus obtained weak convergence results for $U$-statistics in turn extend those obtained by R.G. Miller Jr. and P.K. Sen in [9] in 1972 (cf. Remark 3). The latter results of Miller and Sen are based on the classical condition of the existence of the second moment of the kernel of the underlying $U$-statistic which in turns implies the existence of the second moment of the conditional expected value of the kernel on each of the observations.

*Research supported by a Carleton university Faculty of Graduate Studies and Research scholarship, and NSERC Canada Discovery Grants of M. Csörgő and M. Mojirsheibani at Carleton university.
2 Main results and Background

Let $X_1, X_2, \ldots$ be a sequence of non-degenerate real-valued i.i.d. random variables with distribution $F$. Let $h(X_1, \ldots, X_m)$, symmetric in its arguments, be a Borel-measurable real-valued kernel of order $m \geq 1$, and consider the parameter

$$\theta = \int_{\mathbb{R}^m} \cdots \int h(x_1, \ldots, x_m) \, dF(x_1) \cdots dF(x_m) < \infty.$$ 

The corresponding $U$-statistic (cf. Serfling [10] or Hoeffding [8]) is

$$U_n = \left( \frac{n}{m} \right)^{-1} \sum_{C(n,m)} h(X_{i_1}, \ldots, X_{i_m}),$$

where $m \leq n$ and $\sum_{C(n,m)}$ denotes the sum over $C(n, m) = \{1 \leq i_1 < \ldots < i_m \leq n\}$.

In order to state our results, we first need the following definition.

**Definition.** A sequence $X, X_1, X_2, \ldots$ of i.i.d. random variables is said to be in the domain of attraction of the normal law ($X \in \text{DAN}$) if there exist sequences of constants $A_n$ and $B_n > 0$ such that, as $n \to \infty$,

$$\sum_{i=1}^n X_i - A_n \overset{d}{\to} N(0, 1).$$

**Remark 1.** Further to this definition of $\text{DAN}$, it is known that $A_n$ can be taken as $n \mathbb{E}(X)$ and $B_n = n^{1/2} \ell_X(n)$, where $\ell_X(n)$ is a slowly varying function at infinity (i.e., $\lim_{n \to \infty} \ell_X(nk)/\ell_X(n) = 1$ for any $k > 0$), defined by the distribution of $X$. Moreover, $\ell_X(n) = \sqrt{\text{Var}(X)} > 0$, if $\text{Var}(X) < \infty$, and $\ell_X(n) \to \infty$, as $n \to \infty$, if $\text{Var}(X) = \infty$. Also $X$ has all moments less than 2, and the variance of $X$ is positive, but need not be finite.

Also define the pseudo-self-normalized $U$-process as follows.

$$U_{[nt]}^* = \begin{cases} 0 & , \ 0 \leq t < \frac{m}{n}, \\ \frac{U_{[nt]} - \theta}{V_n} & , \ \frac{m}{n} \leq t \leq 1, \end{cases}$$

where $[.]$ denotes the greatest integer function, $V_n^2 := \sum_{i=1}^n \tilde{h}_i^2(X_i)$ and $\tilde{h}_1(x) = \mathbb{E}(h(X_1, \ldots, X_m) - \theta|X_1 = x)$.

**Theorem 1.** If

(a) $\mathbb{E} \left( |h(X_1, \ldots, X_m)|^{\frac{3}{2}} \log |h(X_1, \ldots, X_m)| \right) < \infty$ and $\tilde{h}_1(X_1) \in \text{DAN},$

then, as $n \to \infty$, we have

(b) $\frac{[nt_0]}{m} U_{[nt_0]}^* \overset{d}{\to} N(0, t_0)$, for $t_0 \in (0, 1]$.
(c) \[ \frac{[nt]}{m} U_{[nt]}^* \to_d W(t) \text{ on } (D[0,1], \rho), \text{ where } \rho \text{ is the sup-norm for functions in } D[0,1] \text{ and } \{W(t), 0 \leq t \leq 1\} \text{ is a standard Wiener process;} \]

(d) On an appropriate probability space for \(X_1, X_2, \ldots,\) we can construct a standard Wiener process \(\{W(t), 0 \leq t < \infty\}\) such that

\[
\sup_{0 \leq t \leq 1} \left| \frac{[nt]}{m} U_{[nt]}^* - \frac{W(nt)}{n^{\frac{1}{2}}} \right| = o_P(1).
\]

**Remark 2.** The statement (c), whose notion will be used throughout, stands for the following functional central limit theorem (cf. Remark 2.1 in Csörgő, Szyszkowicz and Wang [3]). On account of (d), as \(n \to \infty,\) we have

\[
g(S_{[nt]}/V_n) \to_d g(W(.)).
\]

for all \(g : D = D[0,1] \to \mathbb{R}\) that are \((D, \mathcal{D})\) measurable and \(\rho\)-continuous, or \(\rho\)-continuous except at points forming a set of Wiener measure zero on \((D, \mathcal{D}),\) where \(\mathcal{D}\) denotes the \(\sigma\)-field of subsets of \(D\) generated by the finite-dimensional subsets of \(D.\)

Theorem 1 is fashioned after the work on weak convergence of self-normalized partial sums processes of Csörgő, Szyszkowicz and Wang in [2], [3] and [4], which constitute extensions of the contribution of Giné, Götze and Mason in [6].

As to \(\tilde{h}_1(X_1) \in DAN,\) since \(\mathbb{E}\tilde{h}_1(X_1) = 0\) and \(\tilde{h}_1(X_1), \tilde{h}_1(X_2), \ldots,\) are i.i.d. random variables, Theorem 1 of [2] (cf. also Theorem 2.3 of [3]) in this context reads as follows.

**Lemma 1.** As \(n \to \infty,\) the following statements are equivalent:

(a) \(\tilde{h}_1(X_1) \in DAN;\)

(b) \(\frac{\sum_{i=1}^{[nt]} \tilde{h}_1(X_i)}{V_n} \to_d N(0, t_0) \text{ for } t_0 \in (0,1];\)

(c) \(\frac{\sum_{i=1}^{[nt]} \tilde{h}_1(X_i)}{V_n} \to_d W(t) \text{ on } (D[0,1], \rho), \text{ where } \rho \text{ is the sup-norm metric for functions in } D[0,1] \text{ and } \{W(t), 0 \leq t \leq 1\} \text{ is a standard Wiener process;}\)

(d) On an appropriate probability space for \(X_1, X_2, \ldots,\) we can construct a standard Wiener process \(\{W(t), 0 \leq t < \infty\}\) such that

\[
\sup_{0 \leq t \leq 1} \left| \frac{\sum_{i=1}^{[nt]} \tilde{h}_1(X_i)}{V_n} - \frac{W(nt)}{n^{\frac{1}{2}}} \right| = o_P(1).
\]
Also, in the same vein, Proposition 2.1 of [3] for $\tilde{h}_1 (X_1) \in \text{DAN}$ reads as follows.

**Lemma 2.** As $n \to \infty$, the following statements are equivalent:

(a) $\tilde{h}_1 (X_1) \in \text{DAN}$;

There is a sequence of constants $B_n \nearrow \infty$, such that

(b) $\frac{\sum_{i=1}^{[nt_0]} \tilde{h}_1 (X_i)}{B_n} \to_d N(0, t_0)$ for $t_0 \in (0, 1]$;

(c) $\frac{\sum_{i=1}^{[nt]} \tilde{h}_1 (X_i)}{B_n} \to_d W(t)$ on $(D[0, 1], \rho)$, where $\rho$ is the sup-norm metric for functions in $D[0, 1]$ and $\{W(t), 0 \leq t \leq 1\}$ is a standard Wiener process;

(d) On an appropriate probability space for $X_1, X_2, \ldots$, we can construct a standard Wiener process $\{W(t), 0 \leq t < \infty\}$ such that

$$\sup_{0 \leq t \leq 1} \left| \frac{\sum_{i=1}^{[nt]} \tilde{h}_1 (X_i)}{B_n} - \frac{W(nt)}{n^{\frac{1}{2}}} \right| = o_P(1).$$

In view of Lemma 2, a scalar normalized companion of Theorem 1 reads as follows.

**Theorem 2.** If

(a) $E \left( |h(X_1, \ldots, X_m)|^{\frac{3}{2}} \log |h(X_1, \ldots, X_m)| \right) < \infty$ and $\tilde{h}_1 (X_1) \in \text{DAN}$,

then, as $n \to \infty$, we have

(b) $\frac{[nt_0]}{m} \frac{U_{[nt_0]} - \theta}{B_n} \to_d N(0, t_0)$, where $t_0 \in (0, 1]$;

(c) $\frac{[nt]}{m} \frac{U_{[nt]} - \theta}{B_n} \to_d W(t)$ on $(D[0,1], \rho)$, where $\rho$ is the sup-norm for functions in $D[0,1]$ and $\{W(t), 0 \leq t \leq 1\}$ is a standard Wiener process;

(d) On an appropriate probability space for $X_1, X_2, \ldots$, we can construct a standard Wiener process $\{W(t), 0 \leq t < \infty\}$ such that

$$\sup_{0 \leq t \leq 1} \left| \frac{[nt]}{m} \frac{U_{[nt]} - \theta}{B_n} - \frac{W(nt)}{n^{\frac{1}{2}}} \right| = o_P(1).$$
By defining
\[
Y_n^*(t) = 0 \quad \text{for } 0 \leq t \leq \frac{m-1}{n},
\]
\[
Y_n^*(\frac{k}{n}) = \frac{k(U_k - \theta)}{m \sqrt{n \text{Var}(\tilde{h}_1(X_1))}} \quad \text{for } k = m, \ldots, n
\]
and for \( t \in \left[\frac{k-1}{n}, \frac{k}{n}\right], k = m, \ldots, n \),
\[
Y_n^*(t) = Y_n^*(\frac{k-1}{n}) + n(t - \frac{k-1}{n}) \left( Y_n^*(\frac{k}{n}) - Y_n^*(\frac{k-1}{n}) \right),
\]
we can state the already mentioned 1972 weak convergence result of Miller and Sen as follows.

**Theorem A.** If

(I) \( 0 < \mathbb{E}[(h(X_1, X_2, \ldots, X_m) - \theta)(h(X_1, X_{m+1}, \ldots, X_{2m-1}) - \theta)] = \text{Var}(\tilde{h}_1(X_1)) < \infty \)

and

(II) \( \mathbb{E}h^2(X_1, \ldots, X_m) < \infty, \)

then, as \( n \to \infty, \)
\[
Y_n^*(t) \to_d W(t) \quad \text{on } (C[0,1], \rho),
\]
where \( \rho \) is the sup-norm for functions in \( C[0,1] \) and \( \{W(t), 0 \leq t \leq 1\} \) is a standard Wiener process.

**Remark 3.** When \( \mathbb{E}h^2(X_1, \ldots, X_m) < \infty \), first note that existence of the second moment of the kernel \( h(X_1, \ldots, X_m) \) implies the existence of the second moment of \( \tilde{h}_1(X_1) \). Therefore, according to Remark 1, \( B_n = \sqrt{n \mathbb{E}\tilde{h}_1^2(X_1)} \). This means that under the conditions of Theorem A, Theorem 2 holds true and, via (c) of latter, it yields a version of Theorem A on \( D[0,1] \). We note in passing that our method of proofs differs from that of cited paper of Miller and Sen. We use a method of truncation à la \[5\] to relax the condition \( \mathbb{E}h^2(X_1, \ldots, X_m) < \infty \) to the less stringent moment condition \( \mathbb{E} \left( |h(X_1, \ldots, X_m)|^{\frac{4}{3}} \log |h(X_1, \ldots, X_m)| \right) < \infty \) that, in turn, enables us to have \( \tilde{h}_1(X_1) \in \text{DAN} \) in general, with the possibility of infinite variance.

**Remark 4.** Theorem 1 of \[2\] (Theorem 2.3 in \[3\]) as well as Proposition 2.1 of \[3\], continue to hold true in terms of Donskerized partial sums that are elements of \( C[0,1] \). Consequently, the same is true for the above stated Lemmas 1 and 2, concerning \( \tilde{h}_1(X_1) \in \text{DAN} \). This in turn, mutatis mutandis, renders appropriate versions of Theorems 1 and 2 to hold true in \( (C[0,1], \rho) \).
Proof of Theorems 1 and 2.
In view of Lemmas 1 and 2, in order to prove Theorems 1 and 2, we only have to prove the following theorem.

Theorem 3. If \( E \left( |h(X_1, \ldots, X_m)| \right) < \infty \) and \( \bar{h}_1(X_1) \in \text{DAN} \) then, as \( n \to \infty \), we have

\[
\sup_{0 \leq t \leq 1} \left| \frac{[nt]}{m} \sum_{i=1}^{[nt]} \bar{h}_1(X_i) \right| = o_P(1),
\]

and

\[
\sup_{0 \leq t \leq 1} \left| \frac{[nt]}{m} \frac{U_{[nt]} - \bar{h}_1(X_i)}{B_n} \right| = o_P(1).
\]

Proof of Theorem 3. In view of (b) of Lemma 2 with \( t_0 = 1 \), Corollary 2.1 of [3], yields \( V_n B_n \to_P 1 \). This in turn implies the equivalency of (1) and (2). Therefore, it suffices to prove (2) only.

It can be easily seen that

\[
\sup_{0 \leq t \leq 1} \left| \frac{[nt]}{m} \frac{U_{[nt]} - \bar{h}_1(X_i)}{B_n} \right| \leq \sup_{0 \leq t < \frac{m}{n}} \left| \sum_{i=1}^{[nt]} \bar{h}_1(X_i) \right| + \sup_{\frac{m}{n} \leq t \leq 1} \left| \frac{[nt]}{m} \frac{U_{[nt]} - \bar{h}_1(X_i)}{B_n} \right|.
\]

Since, as \( n \to \infty \), we have \( \frac{m}{n} \to 0 \) and, consequently, in view of (d) of Lemma 2

\[
\sup_{0 \leq t < \frac{m}{n}} \left| \sum_{i=1}^{[nt]} \bar{h}_1(X_i) \right| = o_P(1),
\]

in order to prove (2), it will be enough to show that

\[
\sup_{\frac{m}{n} \leq t \leq 1} \left| \frac{[nt]}{m} \frac{U_{[nt]} - \bar{h}_1(X_i)}{B_n} \right| = o_P(1),
\]

or equivalently to show that

\[
\max_{m \leq k \leq n} \left| \frac{k}{mB_n} \left( \frac{k}{m} \right)^{-1} \sum_{C(k,m)} (h(X_{i_1}, \ldots, X_{i_m}) - \bar{h}_1(X_i)) - \frac{1}{B_n} \sum_{i=1}^{k} \bar{h}_1(X_i) \right| = o_P(1).
\]

(3)
We will show that
\[ \sum_{C(k,m)} \left( \bar{h}_1(X_{i_1}) + \ldots + \bar{h}_1(X_{i_m}) \right) = \frac{m}{k} \sum_{i=1}^{k} \bar{h}_1(X_i), \]
where \( \sum_{C(k,m)} \) denotes the sum over \( C(k,m) = \{ 1 \leq i_1 < \ldots < i_m \leq k \} \). To establish (3), without loss of generality we can, and shall assume that \( \theta = 0 \).

Considering that for large \( n \), \( \frac{1}{B_n} \leq \frac{1}{\sqrt{n}} \) (cf. Remark 1), to conclude (3), it will be enough to show that, as \( n \to \infty \), the following holds:

\[ n^{-\frac{1}{2}} \max_{m \leq k \leq n} \left| k \left( \frac{k}{m} \right)^{-1} \sum_{C(k,m)} \left( \bar{h}(X_{i_1}, \ldots, X_{i_m}) - \bar{h}_1(X_{i_1}) - \ldots - \bar{h}_1(X_{i_m}) \right) \right| = o_P(1). \quad (4) \]

To establish (4), for the ease of notation, let

\[ \bar{h}^{(1)}(X_{i_1}, \ldots, X_{i_m}) := h(X_{i_1}, \ldots, X_{i_m})I_{(|h| \leq n^{\frac{3}{2}})} - \mathbb{E}(h(X_{i_1}, \ldots, X_{i_m})I_{(|h| \leq n^{\frac{3}{2}})}), \]
\[ \bar{h}^{(1)}(X_{i_j}) := \mathbb{E}(h^{(1)}(X_{i_1}, \ldots, X_{i_m})|X_{i_j}), \quad j = 1, \ldots, m, \]
\[ \psi^{(1)}(X_{i_1}, \ldots, X_{i_m}) := h^{(1)}(X_{i_1}, \ldots, X_{i_m}) - \bar{h}^{(1)}(X_{i_1}) - \ldots - \bar{h}^{(1)}(X_{i_m}), \]
\[ \bar{h}^{(2)}(X_{i_1}, \ldots, X_{i_m}) := h(X_{i_1}, \ldots, X_{i_m})I_{(|h| > n^{\frac{3}{2}})} - \mathbb{E}(h(X_{i_1}, \ldots, X_{i_m})I_{(|h| > n^{\frac{3}{2}})}), \]
\[ \bar{h}^{(2)}(X_{i_j}) := \mathbb{E}(h^{(2)}(X_{i_1}, \ldots, X_{i_m})|X_{i_j}), \quad j = 1, \ldots, m, \]

where \( I_A \) is the indicator function of the set \( A \). Now observe that

\[ n^{-\frac{1}{2}} \max_{m \leq k \leq n} \left| k \left( \frac{k}{m} \right)^{-1} \sum_{C(k,m)} \left( \bar{h}(X_{i_1}, \ldots, X_{i_m}) - \bar{h}_1(X_{i_1}) - \ldots - \bar{h}_1(X_{i_m}) \right) \right| \]
\[ \leq n^{-\frac{1}{2}} \max_{m \leq k \leq n} \left| k \left( \frac{k}{m} \right)^{-1} \sum_{C(k,m)} \left( h(X_{i_1}, \ldots, X_{i_m}) - h^{(1)}(X_{i_1}, \ldots, X_{i_m}) \right) \right| \]
\[ + n^{-\frac{1}{2}} \max_{m \leq k \leq n} \left| k \left( \frac{k}{m} \right)^{-1} \sum_{C(k,m)} \left( \bar{h}_1(X_{i_1}) + \ldots + \bar{h}_1(X_{i_m}) - \bar{h}^{(1)}(X_{i_1}) - \ldots - \bar{h}^{(1)}(X_{i_m}) \right) \right| \]
\[ + n^{-\frac{1}{2}} \max_{m \leq k \leq n} \left| k \left( \frac{k}{m} \right)^{-1} \sum_{C(k,m)} \psi^{(1)}(X_{i_1}, \ldots, X_{i_m}) \right| \]
\[ := J_1(n) + J_2(n) + J_3(n). \]

We will show that \( J_s(n) = o_P(1), \ s = 1, 2, 3. \)
To deal with the term $J_1(n)$, first note that
\[ h(X_{i_1}, \ldots, X_{i_m}) - h^{(1)}(X_{i_1}, \ldots, X_{i_m}) = h^{(2)}(X_{i_1}, \ldots, X_{i_m}). \]

Therefore, in view of Theorem 2.3.3 of [1] page 43, for $\epsilon > 0$, we can write
\[
P \left( n^{-1} \max_{m \leq k \leq n} k \left( \frac{k}{m} \right)^{-1} \sum_{C(k,m)} h^{(2)}(X_{i_1}, \ldots, X_{i_m}) > \epsilon \right) \\
\leq \epsilon^{-1} n^{-1} \left( m \mathbb{E}|h^{(2)}(X_1, \ldots, X_m)| + n \mathbb{E}|h^{(2)}(X_1, \ldots, X_m)| \right) \\
\leq \epsilon^{-1} n^{-\frac{1}{2}} 2m \mathbb{E}|h(X_1, \ldots, X_m)| + \epsilon^{-1} n^{\frac{1}{2}} 2m \mathbb{E}(|h(X_1, \ldots, X_m)|I_{(|h|>n^{\frac{1}{2}})}) \\
\leq \epsilon^{-1} n^{-\frac{1}{2}} 2m \mathbb{E}|h(X_1, \ldots, X_m)| + \epsilon^{-1} 2m \mathbb{E}(|h(X_1, \ldots, X_m)|I_{(|h|>n^{\frac{1}{2}})}) \\
\to 0, \quad \text{as } n \to \infty.
\]

Here we have used the fact that $\mathbb{E}|h(X_1, \ldots, X_m)|^{\frac{3}{2}} < \infty$. The last line above implies that $J_1(n) = o_P(1)$.

Next to deal with $J_2(n)$, first observe that
\[ \tilde{h}_1(X_{i_1}) + \ldots + \tilde{h}_1(X_{i_m}) - \tilde{h}^{(1)}(X_{i_1}) - \ldots - \tilde{h}^{(1)}(X_{i_m}) = \sum_{j=1}^m \tilde{h}^{(2)}(X_{i_j}). \]

It can be easily seen that $\sum_{j=1}^m \tilde{h}^{(2)}(X_{i_j})$ is symmetric in $X_{i_1}, \ldots, X_{i_m}$. Thus, in view of Theorem 2.3.3 of [1] page 43, for $\epsilon > 0$, we have
\[
P \left( n^{-1} \max_{m \leq k \leq n} k \left( \frac{k}{m} \right)^{-1} \left( \sum_{j=1}^m \tilde{h}^{(2)}(X_{i_j}) \right) > \epsilon \right) \\
\leq \epsilon^{-1} n^{-\frac{1}{2}} 2m \mathbb{E}|h(X_1, \ldots, X_m)| + \epsilon^{-1} n^{\frac{1}{2}} 2m \mathbb{E}(|h(X_1, \ldots, X_m)|I_{(|h|>n^{\frac{1}{2}})}) \\
\to 0, \quad \text{as } n \to \infty,
\]
i.e., $J_2(n) = o_P(1)$.

**Note.** Alternatively, one can use Etemadi’s maximal inequality for partial sums of i.i.d. random variables, followed by Markov inequality, to show $J_2(n) = o_P(1)$. 

8
As for the term $J_3(n)$, first note that $\binom{k}{m}^{-1} \sum_{C(k,m)} \psi^{(1)}(X_{i_1}, \ldots, X_{i_m})$ is a $U$-statistic. Consequently one more application of Theorem 2.3.3 page 43 of [1] yields,

$$
\Pr \left( n^{\frac{1}{2}} \max_{m \leq k \leq n} \left( \binom{k}{m}^{-1} \sum_{C(k,m)} \psi^{(1)}(X_{i_1}, \ldots, X_{i_m}) > \epsilon \right) \right) 
\leq n^{-1} \epsilon^{-2} m^2 \mathbb{E}(\psi^{(1)}(X_1, \ldots, X_m))^2 
+ n^{-1} \epsilon^{-2} \sum_{k=m+1}^{n} (2k+1) \mathbb{E} \left( \binom{k}{m}^{-1} \sum_{C(k,m)} \psi^{(1)}(X_{i_1}, \ldots, X_{i_m}) \right)^2. 
$$

(5)

Observing that $\mathbb{E}(\psi^{(1)}(X_1, \ldots, X_m))^2 \leq C(m) \mathbb{E} \left( h^2(X_1, \ldots, X_m)I_{(|h| \leq n^{\frac{3}{2}})} \right)$, where $C(m)$ is a positive constant that does not depend on $n$,

$$
\mathbb{E}\psi^{(1)}(X_1, \ldots, X_m) = \mathbb{E}(\psi^{(1)}(X_1, \ldots, X_m)|X_j = 0, j = 1, \ldots, m,
$$

and in view of Lemma B page 184 of [10], it follows that for some positive constants $C_1(m)$ and $C_2(m)$ which do not depend on $n$, the R.H.S. of (5) is bounded above by

$$
\epsilon^{-2} n^{-1} \mathbb{E} \left( h^2(X_1, \ldots, X_m)I_{(|h| \leq n^{\frac{3}{2}})} \right) (C_1(m) + C_2(m) \log(n)) 
\leq \epsilon^{-2} C_1(m) n^{\frac{1}{2}} \mathbb{E}|h(X_1, \ldots, X_m)|^{\frac{4}{3}} 
+ \epsilon^{-2} C_1(m) \mathbb{E} \left( |h(X_1, \ldots, X_m)|^{\frac{4}{3}} I_{(n<|h| \leq n^{\frac{3}{2}})} \right) 
+ \epsilon^{-2} C_2(m) n^{\frac{1}{2}} \log(n) \mathbb{E}|h(X_1, \ldots, X_m)|^{\frac{4}{3}} 
+ \epsilon^{-2} C_2(m) \mathbb{E} \left( |h(X_1, \ldots, X_m)|^{\frac{4}{3}} \log|h(X_1, \ldots, X_m)| \right) I_{(n<|h| \leq n^{\frac{3}{2}})} 
\leq \epsilon^{-2} C_1(m) n^{\frac{1}{2}} \mathbb{E}|h(X_1, \ldots, X_m)|^{\frac{4}{3}} 
+ \epsilon^{-2} C_1(m) \mathbb{E} \left( |h(X_1, \ldots, X_m)|^{\frac{4}{3}} I_{(|h|>n)} \right) 
+ \epsilon^{-2} C_2(m) n^{\frac{1}{2}} \log(n) \mathbb{E}|h(X_1, \ldots, X_m)|^{\frac{4}{3}} 
+ \epsilon^{-2} C_2(m) \mathbb{E} \left( |h(X_1, \ldots, X_m)|^{\frac{4}{3}} \log|h(X_1, \ldots, X_m)| \right) I_{(|h|>n)} 
\rightarrow 0, \text{ as } n \rightarrow \infty.
$$

Thus $J_3(n) = o_P(1)$. This also completes the proof of (4), and hence also that of Theorem 3. Now, as already noted above, the proof of Theorems 1 and 2 follow from Theorem 3 and Lemmas 1 and 2.

**Remark 5.** Studying a $U$-statistics type process that can be written as a sum of three $U$-statistics of order $m = 2$, Csörgő, Szyszkowicz and Wang in [5] proved that under the slightly more relaxed condition that $\mathbb{E}|h(X_1, \ldots, X_m)|^{\frac{4}{3}} < \infty,$
as \( n \to \infty \), we have
\[
n^{-3/2} \max_{1 \leq k \leq n} \sum_{1 \leq i < j \leq k} (h(X_i, X_j) - \tilde{h}_1(X_i) - \tilde{h}_1(X_j)) = o_P(1).
\]

In the proof of the latter, the well known Doob maximal inequality for martingales was used, which gives us a sharper bound. The just mentioned inequality is not applicable for the processes in Theorems 1 and 2, even for \( U \)-statistics of order 2. The reason for this is that the inside parts of the absolute values of \( J_s(n) \), \( s = 1, 2, 3 \), are not martingales. Also, since \( \sum_{C(k,m)} (h(X_{i_1}, \ldots, X_{i_m}) - \tilde{h}_1(X_{i_1}) - \ldots - \tilde{h}_1(X_{i_m})) \), for \( m > 2 \), no longer form a martingale, it seems that the Doob maximal inequality is not applicable for the process
\[
n^{-m+\frac{1}{2}} \max_{1 \leq k \leq n} \sum_{C(k,m)} (h(X_{i_1}, \ldots, X_{i_m}) - \tilde{h}_1(X_{i_1}) - \ldots - \tilde{h}_1(X_{i_m})),
\]
which is an extension of the \( U \)-statistics parts of the process used by Csörgő, Szyszkowicz and Wang in [5] for \( m = 2 \).

Due to the nonexistence of the second moment of the kernel of the underlying \( U \)-statistic in the following example, the weak convergence result of Theorem A fails to apply. However, using Theorem 1 for example, one can still derive weak convergence results for the underlying \( U \)-statistic.

**Example.** Let \( X_1, X_2, \ldots \), be a sequence of i.i.d. random variables with the density function
\[
f(x) = \begin{cases} |x - a|^{-3}, & |x - a| \geq 1, \ a \neq 0, \\ 0, & \text{elsewhere}. \end{cases}
\]
Consider the parameter \( \theta = \mathbb{E}^m(X_1) = a^m \), where \( m \geq 1 \) is a positive integer, and the kernel \( h(X_1, \ldots, X_m) = \prod_{i=1}^m X_i \). Then with \( m, n \) satisfying \( n \geq m \), the corresponding \( U \)-statistic is
\[
U_n = \left( \frac{n}{m} \right)^{-1} \sum_{C(n,m)} \prod_{j=1}^m X_{i_j}.
\]
Simple calculation shows that \( \tilde{h}_1(X_1) = X_1 a^{m-1} - a^m \).

It is easy to check that \( \mathbb{E} \left( |h(X_1, \ldots, X_m)|^{\frac{4}{3}} \log |h(X_1, \ldots, X_m)| \right) < \infty \) and that \( \tilde{h}_1(X_1) \in \text{DAN} \) (cf. Gut [7], page 439). In order to apply Theorem 1 for this \( U \)-statistic, define
\[
U^*_\left[ nt \right] = \begin{cases} 0, & 0 \leq t < \frac{m}{n}, \\ \left( \frac{n}{m} \right)^{-1} \sum_{C(\left[ nt \right],m)} \prod_{i=1}^m X_{i_j} - a^m \left( \sum_{i=1}^n (X_i a^{m-1} - a^m)^2 \right)^{\frac{1}{2}}, & \frac{m}{n} \leq t \leq 1. \end{cases}
\]
Then, based on (c) of Theorem 1, as \( n \to \infty \), we have

\[
\frac{[nt]}{m} U^*_n \longrightarrow d W(t) \text{ on } (D[0,1], \rho),
\]

where \( \rho \) is the sup-norm metric for functions in \( D[0,1] \) and \( \{W(t), \ 0 \leq t \leq 1\} \) is a standard Wiener process. Taking \( t = 1 \) gives us a central limit theorem for the pseudo-self-normalized \( U \)-statistic

\[
U^*_n = \frac{(\frac{n}{m})^{-1} \sum_{C(n,m)} \prod_{j=1}^{m} X_{ij} - a^m}{\left( \sum_{i=1}^{n} (X_i a^{m-1} - a^m)^2 \right)^{\frac{1}{2}}}.
\]

i.e., as \( n \to \infty \), we have

\[
\frac{n}{m} U^*_n \longrightarrow d N(0,1).
\]

Acknowledgments. The author wishes to thank Miklós Csörgő, Barbara Szyszkowicz and Qiying Wang for calling his attention to a preliminary version of their paper [5] that inspired the truncation arguments of the present exposition. This work constitutes a part of the author’s Ph.D. thesis in preparation, written under the supervision and guidance of Miklós Csörgő and Majid Mojirsheibani. My special thanks to them for also reading preliminary versions of this article, and for their instructive comments and suggestions that have much improved the construction and presentation of the results.
References

[1] Borovskikh, Yu. V. (1996). *U-statistics in Banach Spaces*. VSP, Utrecht.

[2] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2003). Donsker’s theorem for self-normalized parial sums processes. *The Annals of Probability* **31**, 1228-1240.

[3] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2004). On Weighted Approximations and Strong Limit Theorems for Self-normalized Partial Sums Processes. In *Asymptotic methods in Stochastics*, 489-521, Fields Inst. Commun. **44**, Amer. Math. Soc., Providence, RI.

[4] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2008). On weighted approximations in $D[0, 1]$ with application to self-normalized partial sum processes. *Acta Mathematica Hungarica* **121** (4), 307-332.

[5] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2008). Asymptotics of studentized U-type processes for changepoint problems. *Acta Mathematica Hungarica* **121** (4), 333-357.

[6] Giné, E., Götze, F. and Mason D. M. (1997). When is the student t-statistic asymptotically Normal? *The Annals of Probability* **25**, 1514-1531.

[7] Gut, A. (2005). *Probability: A Graduate Course*. Springer.

[8] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.

[9] Miller, R. G. Jr. and Sen, P. K. (1972). Weak convergence of U-statistics and Von Mises’ differentiable statistical functions. *Ann. Math. Statist.* **43**, 31-41.

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