Law of large numbers for Hilbert valued U-statistics of mixing data

Davide Giraudo (Université de Strasbourg, France)

Monday, 12 August 2024,
11th World Congress in Probability and Statistics
Bochum

Plan

Definition of U-statistics

In order to estimate parameters expressable as $\mathbb{E}\left[h\left(X, Y\right)\right]$ via an empirical mean, where X and Y are i.i.d. and $h\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, the U-statistic of kernel h, defined

$$U_n := \sum_{1 \leqslant i < j \leqslant n} h\left(X_i, X_j\right), \quad n \geqslant 2,$$

where $\left(X_j\right)_{j \geqslant 1}$ is i.i.d., was introduced by **Hoeffding** (1948).

One can take as estimator $U_n / \binom{n}{2}$, which is unbiaised.

General goal : understand the asymptotic behavior of $\left(U_n\right)_{n \geqslant 2}$.

Notice that for each $i < j$, $h\left(X_i, X_j\right)$ has the same law as $h\left(X_1, X_2\right)$. The U-statistic U_n can be viewed as partial sums of the non-independent random variables $D_j := \sum_{i=1}^{j-1} h\left(X_i, X_j\right)$.

Examples

1. Suppose that $(X_i)_{i\geqslant 1}$ is i.i.d., centered, and such that $\mathbb{E}\left[|X_1|^p\right] < \infty$ for some $1 \leqslant p < 2$ and $h(x, y) = x + y$. Then

$$U_n = (n-1)\sum_{k=1}^{n} X_k$$

   hence $U_n/n^{1+1/p} \to 0$ almost surely.

2. Suppose that $(X_i)_{i\geqslant 1}$ is i.i.d., centered, $\mathbb{E}\left[|X_1|^p\right] < \infty$ for some $1 \leqslant p < 2$ and $h(x, y) = x \cdot y$. Then

$$U_n = \frac{1}{2}\left(\left(\sum_{k=1}^{n} X_k\right)^2 - \sum_{k=1}^{n} X_k^2\right)$$

   hence $U_n/n^{2/p} \to 0$ almost surely for each $1 \leqslant p < 2$. In this case, the weaker normalization $n^{2/p}$ can be taken.

Other examples of kernels

Recall that

$$U_n := \sum_{1 \leqslant i < j \leqslant n} h(X_i, X_j), \quad n \geqslant 2.$$

3. Variance estimator : $h(x, y) := (x - y)^2 / 2$.
4. Gini mean differences : $h(x, y) = |x - y|$.
5. Grassberger-Procaccia estimator : for fixed $t > 0$, $h(x, y) = \mathbf{1}\{|x - y| \leqslant t\}$.
6. $h(x, y) = \operatorname{sgn}(x - y)$.

Martingale property ?

Let $D_j := \sum_{i=1}^{j-1} h(X_i, X_j)$ where $(X_i)_{i\geqslant 0}$ is i.i.d.. Then $U_n = \sum_{j=2}^{n} D_j$. We would like to know whether $(D_j)_{j\geqslant 2}$ is a martingale difference sequence for the filtration $(\mathcal{F}_j)_{j\geqslant 1}$, where $\mathcal{F}_j = \sigma(X_k, 1 \leqslant k \leqslant j)$.
Using the property

$$\mathbb{E}[Y \mid \mathcal{F} \vee \mathcal{G}] = \mathbb{E}[Y \mid \mathcal{F}]$$

valid if $\mathcal{G}$ is independent of $\sigma(Y) \vee \mathcal{F}$, we derive

$$\begin{aligned}
\mathbb{E}[D_j \mid \mathcal{F}_{j-1}] &= \sum_{i=1}^{j-1} \mathbb{E}[h(X_i, X_j) \mid \sigma(X_k, 1 \leqslant k \leqslant j-1)] \\
&= \sum_{i=1}^{j-1} \mathbb{E}[h(X_i, X_j) \mid \sigma(X_i) \vee \sigma(X_k, 1 \leqslant k \leqslant j-1, k \neq i)] \\
&= \sum_{i=1}^{j-1} \mathbb{E}[h(X_i, X_j) \mid X_i] \\
&= \sum_{i=1}^{j-1} h_1(X_i), \text{ with } h_1(x) = \mathbb{E}[h(x, X_2)]
\end{aligned}$$

hence $\mathbb{E}[D_j \mid \mathcal{F}_{j-1}] = 0$ if and only if $\mathbb{E}[h(X_1, X_2) \mid X_1] = 0$.

Tool : Hoeffding's decomposition

Let $h\colon \mathbb{R}^2 \to \mathbb{R}$ be a measurable function and let $(X_i)_{i\geqslant 1}$ be an i.i.d. sequence. Let

$$U_n = \sum_{1\leqslant i<j\leqslant n} h\left(X_i, X_j\right).$$

We define $\theta := \mathbb{E}\left[h\left(X_1, X_2\right)\right]$,

$$h_1\left(x\right) = \mathbb{E}\left[h\left(x, X_2\right)\right] - \theta, \quad h_2\left(y\right) = \mathbb{E}\left[h\left(X_1, y\right)\right] - \theta,$$

$$h_3\left(x, y\right) = h\left(x, y\right) - h_1\left(x\right) - h_2\left(y\right) - \theta.$$

Then

$$U_n = \binom{n}{2}\theta + \sum_{i=1}^{n-1}\left(n-i\right)h_1\left(X_i\right) + \sum_{j=2}^{n}\left(j-1\right)h_2\left(X_j\right) + \sum_{1\leqslant i<j\leqslant n} h_3\left(X_i, X_j\right)$$

and

$$\mathbb{E}\left[h_3\left(X_i, X_j\right) \mid X_1, \ldots, X_{j-1}\right] = \mathbb{E}\left[h_3\left(X_i, X_j\right) \mid X_{i+1}, \ldots, X_n\right] = 0.$$

Symmetric case

We assume in this slide that h is symmetric, that is, $h(x,y) = h(y,x)$ for each $x, y \in \mathbb{R}$. We got previously

$$U_n = \binom{n}{2}\theta + \sum_{i=1}^{n-1}(n-i)\,h_1(X_i) + \sum_{j=2}^{n}(j-1)\,h_2(X_j) + \sum_{1\leqslant i<j\leqslant n} h_3(X_i, X_j)\,.$$

Symmetric of h implies that $h_1 = h_2$ hence

$$U_n = \binom{n}{2}\theta + (n-1)\sum_{i=1}^{n} h_1(X_i) + \sum_{1\leqslant i<j\leqslant n} h_3(X_i, X_j)\,.$$

The term $(n-1)\sum_{i=1}^{n} h_1(X_i)$ is called linear part ; the term $\sum_{1\leqslant i<j\leqslant n} h_3(X_i, X_j)$ degenerate part.

We say that h is degenerate if $\mathbb{E}\left[h(X_1, X_2) \mid X_1\right] = 0$ a.s.

When h is not supposed to be symmetric, degeneracy means $\mathbb{E}\left[h(X_1, X_2) \mid X_1\right] = \mathbb{E}\left[h(X_1, X_2) \mid X_2\right] = 0$.

Treatment of the degenerated part

Denote $U_n(h_3) = \sum_{1\leqslant i<j\leqslant n} h_3(X_i, X_j)$ the degenerate part.
We combine the martingale property for the summation over i and j with Burkholder's inequality.

Moment of order 2 **:** if $(i,j) \neq (k,\ell)$, then $\mathbb{E}[h_3(X_i,X_j)h_3(X_k,X_\ell)] = 0$ (indeed, if $j \neq \ell$, one has

$$\mathbb{E}[h_3(X_i,X_j)h_3(X_k,X_\ell)] = \mathbb{E}\left[\mathbb{E}\left[h_3(X_i,X_j)h_3(X_k,X_\ell) \mid X_1,\dots,X_{\min\{j,\ell\}-1}\right]\right] = 0$$

and if $j = \ell$, then necessarily $i \neq k$ and

$$\mathbb{E}[h_3(X_i,X_j)h_3(X_k,X_\ell)] = \mathbb{E}\left[\mathbb{E}\left[h_3(X_i,X_j)h_3(X_k,X_\ell) \mid X_{\max\{i,k\}+1},\dots,X_\ell\right]\right] = 0).$$

Consequently,

$$\mathbb{E}\left[\max_{2\leqslant n\leqslant N} U_n(h_3)^2\right] \leqslant 4\sum_{1\leqslant i<j\leqslant N}\mathbb{E}\left[h_3^2(X_i,X_j)\right] \leqslant KN^2\mathbb{E}\left[h_3^2(X_1,X_2)\right].$$

Moment of order $1 < p < 2$ **:**

$$\mathbb{E}\left[\max_{2\leqslant n\leqslant N}|U_n(h_3)|^p\right] \leqslant K_p N^2\mathbb{E}[|h_3(X_1,X_2)|^p].$$

Moment of order $p > 2$ **:**

$$\mathbb{E}\left[\max_{2\leqslant n\leqslant N}|U_n(h_3)|^p\right] \leqslant K_p N^p\mathbb{E}[|h_3(X_1,X_2)|^p].$$

Law of large numbers

Hoeffding (1948) showed that if $(X_i)_{i \geqslant 1}$ is i.i.d. and $\mathbb{E}\left[\left|h\left(X_1, X_2\right)\right|\right] < \infty$, then $U_n(h) / \binom{n}{2} \to \mathbb{E}\left[h\left(X_1, X_2\right)\right]$ almost surely.

Proposition (**Giné, Zinn** (1991))

Let $(X_i)_{i \geqslant 1}$ be an i.i.d. sequence and $h \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a measurable function. Let $1 \leqslant p < 2$. Suppose that $\mathbb{E}\left[\left|h\left(X_1, X_2\right)\right|^p\right] < \infty$.

If h is degenerate with respect to $(X_i)_{i \geqslant 1}$ (that is, $\mathbb{E}\left[h\left(X_1, X_2\right) \mid X_1\right] = 0 = \mathbb{E}\left[h\left(X_1, X_2\right) \mid X_2\right]$ a.s.), then

$$\frac{1}{n^{2/p}} \left| \sum_{1 \leqslant i < j \leqslant n} h\left(X_i, X_j\right) \right| \to 0 \text{ a.s.}$$

If we only assume that $\mathbb{E}\left[h\left(X_1, X_2\right)\right] = 0$, then

$$\frac{1}{n^{1+1/p}} \left| \sum_{1 \leqslant i < j \leqslant n} h\left(X_i, X_j\right) \right| \to 0 \text{ a.s.}$$

Control of the maximal function

Proposition (G. (2024))

Let $(X_i)_{i\geqslant 1}$ be an i.i.d. sequence and let $h\colon \mathbb{R}\times\mathbb{R}\to\mathbb{H}$ be a measurable function, where $(\mathbb{H},\langle\cdot,\cdot\rangle)$ is a separable Hilbert space. Let $1\leqslant p<2$.

- *If h is degenerate for $(X_i)_{i\geqslant 1}$ (that is, $\mathbb{E}\left[h\left(X_1,X_2\right)\mid X_1\right]=\mathbb{E}\left[h\left(X_1,X_2\right)\mid X_2\right]=0$ a.s.), then*

$$\sup_{t>0} t^p\mathbb{P}\left(\sup_{n\geqslant 1}\frac{1}{n^{2/p}}\left\|\sum_{1\leqslant i<j\leqslant n} h\left(X_i,X_j\right)\right\|>t\right)_{\mathbb{H}}\leqslant \kappa_p\mathbb{E}\left[\left|h\left(X_1,X_2\right)\right|^p\right].$$

- *If we only assume that $\mathbb{E}\left[h\left(X_1,X_2\right)\right]=0$, then*

$$\sup_{t>0} t^p\mathbb{P}\left(\sup_{n\geqslant 1}\frac{1}{n^{1+1/p}}\left\|\sum_{1\leqslant i<j\leqslant n} h\left(X_i,X_j\right)\right\|_{\mathbb{H}}>t\right)\leqslant \kappa_p\mathbb{E}\left[\left|h\left(X_1,X_2\right)\right|^p\right].$$

Generalizations of U-statistic of order two

1. We can consider U-statistics of higher order : for $h\colon \mathbb{R}^m \to \mathbb{R}$,

$$U_n = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant m} h\left(X_{i_1}, \ldots, X_{i_m}\right).$$

2. It is also possible to replace h by a function depending on $(i_1, \ldots, i_m)$:

$$U_n = \sum_{1 \leqslant i_1 < \cdots < i_m \leqslant n} h_{i_1, \ldots, i_m}\left(X_{i_1}, \ldots, X_{i_m}\right).$$

3. The random variables X_i can take their values in a measurable space $(S, \mathcal{S})$.

Plan

Goal

Let $h\colon S^2 \to \mathbb{H}$ be a measurable function, where (S, d) is a separable metric space and $\mathbb{H}$ a separable Hilbert space. Let $(X_i)_{i \geqslant 1}$ be a strictly stationary sequence.

For $1 \leqslant p < 2$, we study the almost sure convergence of

$$\frac{1}{n^{1+1/p}} \left\| \sum_{1 \leqslant i < j \leqslant n} \left(h\left(X_i, X_j\right) - \mathbb{E}\left[h\left(X_i, X_j\right)\right] \right) \right\|_{\mathbb{H}}$$

to 0 and when it is possible, that of

$$\frac{1}{n^{2/p}} \left\| \sum_{1 \leqslant i < j \leqslant n} \left(h\left(X_i, X_j\right) - \mathbb{E}\left[h\left(X_i, X_j\right)\right] \right) \right\|_{\mathbb{H}} .$$

The assumptions will concern the dependence of $(X_i)_{i \geqslant 1}$ and the moments of the random variables $\left\| h\left(X_1, X_j\right) \right\|_{\mathbb{H}}$, $j \geqslant 2$.

Why do we consider Hilbert-space valued kernel and data with values in a metric space ?

Some robust tests (**Chakraborty and Chaudhuri** (2015, 2017) ; **Wegner and Wendler** (2023), **Jiang, Wang and Shao** (2023)) are based on a generalization of the real-valued kernel $h(x, y) = \operatorname{sgn}(x - y)$, which is given by

$$h\colon \mathbb{H} \times \mathbb{H} \to \mathbb{H}, \quad h(x, y) = \begin{cases} \frac{x-y}{\|x-y\|_{\mathbb{H}}} & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

Working with metric space valued data allows to consider for instance functional data.

Mixing coefficients

Given a strictly stationary sequence $(X_i)_{i\geqslant 1}$, we define

$$\beta(k) := \sup_{m\geqslant 1} \beta\left(\sigma\left(X_i, 1\leqslant i\leqslant m\right), \sigma\left(X_i, i\geqslant m+k\right)\right),$$

where

$$\beta(\mathcal{A},\mathcal{B}) := \frac{1}{2}\sup\left\{\sum_{i=1}^{I}\sum_{j=1}^{J}\left|\mathbb{P}\left(A_i\cap B_j\right)-\mathbb{P}\left(A_i\right)\mathbb{P}\left(B_j\right)\right|\right\},$$

and the supremum is taken over finite partitions $(A_i)_{i=1}^{I}, A_i\in\mathcal{A}$ and $(B_j)_{j=1}^{J}, B_j\in\mathcal{B}$ of Ω.

See **Rio** (2000) for some examples of mixing sequences

Approach (1)

One can still do the Hoeffding's decomposition, but the martingale property of the degenerate part does not hold. The convergence of the linear part is guaranted by existing results : **Dedecker et Merlevède** (2003, 2006).

We are reduced to show that for each $\varepsilon > 0$,

$$\sum_{M=0}^{\infty} \mathbb{P}\left(2^{-M\left(1+\frac{1}{p}\right)} \max_{2\leqslant n\leqslant 2^M} \left\| \sum_{1\leqslant i<j\leqslant n} h_3\left(X_i, X_j\right) \right\|_{\mathbb{H}} > \varepsilon\right) < \infty$$

in the non-degenerate case ; in the degenerate one, that is, when

$$\mathbb{E}\left[h\left(X_1, X_1'\right) \mid X_1\right] = \mathbb{E}\left[h\left(X_1, X_1'\right) \mid X_1'\right] = 0$$

(X_1' is an independent copy of X_1), the exponent $1 + 1/p$ has to be replaced by $2/p$.

We thus need to control $\mathbb{P}\left(\max_{2\leqslant n\leqslant 2^M} \left\|\sum_{1\leqslant i<j\leqslant n} h_3\left(X_i, X_j\right)\right\|_{\mathbb{H}} > x\right)$ for $x > 0$.

Approach (2)

For fixed $q \in \left\{1, \dots, 2^{M-1}\right\}$, we express $\sum_{1 \leqslant i<j \leqslant n} h_3\left(X_i, X_j\right)$ as a U-statistic based on the vectors

$$V_{k,u} := \left(X_{2uq+k+1}, \dots, X_{2qu+k+q+1}\right), \quad -q \leqslant k \leqslant q,$$

plus some remainder terms.

For fixed k, by **Berbee** (1979) we can find vectors $V^*_{k,u}$ such that

- for each $u \geqslant 1$, $V_{k,u}$ has the same law as $V^*_{k,u}$;
- $\mathbb{P}\left(V_{k,u} \neq V^*_{k,u}\right) \leqslant \beta\left(q\right)$ and
- $\left(V^*_{k,u}\right)_{u \geqslant 1}$ is independent.

Letting $H = \left\|h\left(X_1, X_1'\right)\right\|_{\mathbb{H}}$, we get, for $r \geqslant 2$, $R, x > 0$ and $1 \leqslant q \leqslant 2^{M-1}$:

$$\mathbb{P}\left(\max_{2 \leqslant n \leqslant 2^M}\left\|\sum_{1 \leqslant i<j \leqslant n} h_3\left(X_i, X_j\right)\right\|_{\mathbb{H}} > x\right) \leqslant C_r x^{-r} q^r 2^{Mr} \mathbb{E}\left[H^r \mathbf{1}_{H \leqslant R}\right]$$
$$+ C_r x^{-1} 2^{2M} \mathbb{E}\left[H \mathbf{1}_{H > R}\right] + C_r x^{-1} q N \sup_{j \geqslant 2} \mathbb{E}\left[\left\|h\left(X_1, X_j\right)\right\|_{\mathbb{H}}\right] + 2^{M+2} \beta\left(q\right).$$

The results for moments of order smaller than two

Recall that

$$U_n(h) = \sum_{1 \leqslant i < j \leqslant n} h(X_i, X_j),$$

where $(X_i)_{i \in \mathbb{Z}}$ is strictly stationary and let $U_n^c = U_n(h) - \mathbb{E}[U_n(h)]$. Let $H := \|h(X_1, X_1')\|_{\mathbb{H}}$, where X_1' is an independent copy of X_1. The required assumptions are of the form

$C_H(q)$: $H \in \mathbb{L}^q$

$C_\beta(\gamma)$: $\sum_{k=1}^{\infty} k^\gamma \beta(k) < \infty$.

We assume that $\sup_{j \geqslant 2} \mathbb{E}\left[\|h(X_1, X_j)\|_{\mathbb{H}}\right] < \infty$. Let $1 < p < 2$ and

$$\gamma(p, \delta) = \max\left\{p - 2 + \frac{p(p-1)}{\delta}, \frac{p(p-1) + (p-1)\delta}{p(p-1) + (p+1)\delta}\right\}.$$

We assume that for some $\delta \in (0, 2-p)$,

Theorem G. (2024)	Non-degenerate case	Degenerate case
Assumption on H :	$C_H(p+\delta)$	$C_H(p+\delta)$
Assumption on $\beta(\cdot)$:	$C_\beta(\gamma(p,\delta))$	$C_\beta\left((p-1)\left(1+\frac{p}{\delta}\right)\right)$
Convergence	$\frac{1}{n^{1+1/p}}\|U_n^c(h)\|_{\mathbb{H}} \to 0$ a.s.	$\frac{1}{n^{2/p}}\|U_n^c(h)\|_{\mathbb{H}} \to 0$ a.s.

The results for moments of order higher than two

Recall that

$$U_n(h) = \sum_{1 \leqslant i < j \leqslant n} h(X_i, X_j),$$

where $(X_i)_{i \in \mathbb{Z}}$ is strictly stationary and let $U_n^c = U_n(h) - \mathbb{E}[U_n(h)]$. Let $H := \|h(X_1, X_1')\|_{\mathbb{H}}$, where X_1' is an independent copy of X_1. The required assumptions are of the form

$C_H(q)$: $H \in \mathbb{L}^q$

$C_\beta(\gamma)$: $\sum_{k=1}^{\infty} k^\gamma \beta(k) < \infty$.

We assume that $\sup_{j \geqslant 2} \mathbb{E}\left[\|h(X_1, X_j)\|_{\mathbb{H}}\right] < \infty$. Let $1 < p < 2$.

Theorem G. (2024)	Non-degenerate case	Degenerate case
Assumption on H	$C_H(p+\delta)$ for some $\delta \geqslant 2-p$	$C_H(2)$
Assumption on $\beta(\cdot)$: for some $\eta > 0$	$C_\beta(p-1+\eta)$ and $C_\beta\left(\frac{p(p-1)}{\delta}\right)$ for the same δ as in C_H	$C_\beta\left(\frac{2(p-1)}{2-p} + \eta\right)$
Convergence	$\frac{1}{n^{1+1/p}} \|U_n^c(h)\|_{\mathbb{H}} \to 0$ a.s.	$\frac{1}{n^{2/p}} \|U_n^c(h)\|_{\mathbb{H}} \to 0$ a.s.

Comparison with previous results

Non-degenerated case : **Dehling, Sharipov** (2009) : with

$$\gamma(p,\delta) = \max\left\{ p - 2 + \frac{p(p-1)}{\delta}, \frac{p(p-1) + (p-1)\delta}{p(p-1) + (p+1)\delta} \right\}$$

replaced by $\max\{p - 2 + p(p-1)/\delta, 1\}$, $\mathbb{H} = \mathbb{R}$ and h symmetric.

Our result also extends that of **Dehling, Sharipov** (2009) in the degenerate case, since we treat the not-necessarily symmetric case and the Hilbert-space valued case.

Remaining questions

1. Treatment of incomplete U-statistics :

$$U_n^{\text{inc}} = \sum_{1 \leqslant i < j \leqslant n} Z_{n,i,j} h(X_i, X_j),$$

   where $(Z_{n,i,j})_{1 \leqslant i < j}$ is i.i.d., independent of $(X_i)_{i \geqslant 1}$ and $Z_{n,i,j}$ has a Bernoulli distribution with parameter p_n.

2. Treatment of U-statistics of higher order : for $h \colon S^m \to \mathbb{H}$,

$$U_n = \sum_{1 \leqslant i_1 < \cdots < i_m \leqslant n} h(X_{i_1}, \ldots, X_{i_m}).$$

3. α-mixing sequences can also be considered. Assumptions on the kernel have to be made, since we can only control the probability that distance between the original random variables and the coupled ones is bigger than some number.