



# **An exponential inequality for Hilbert-valued U-statistics of i.i.d. data**

Davide Giraud

## **► To cite this version:**

| Davide Giraud. An exponential inequality for Hilbert-valued U-statistics of i.i.d. data. 2024. <hal-04694894>

**HAL Id: hal-04694894**

**<https://hal.science/hal-04694894v1>**

Preprint submitted on 12 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# An exponential inequality for Hilbert-valued $U$ -statistics of i.i.d. data

Davide Giraudo<sup>1</sup>

<sup>†</sup>Institut de Recherche Mathématique Avancée UMR 7501, Université de Strasbourg and CNRS 7 rue René Descartes 67000 Strasbourg, France

September 12, 2024

## Abstract

In this paper, we establish an exponential inequality for  $U$ -statistics of i.i.d. data, varying kernel and taking values in a separable Hilbert space. The bound are expressed as a sum of an exponential term plus an other one involving the tail of a sum of squared norms. We start by the degenerate case. Then we provide applications to  $U$ -statistics of not necessarily degenerate fixed kernel, weighted  $U$ -statistics and incomplete  $U$ -statistics.

## 1 An exponential inequality for Hilbert-valued degenerate $U$ -statistics

### 1.1 Definition of $U$ -statistics

Given an i.i.d. sequence  $(\xi_i)_{i \geq 1}$  taking values in a measurable space  $(S, \mathcal{S})$ , a separable Banach space  $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$  and measurable functions  $h_{\mathbf{i}}: S^m \rightarrow \mathbb{B}$  (where  $S^m$  is endowed with the product  $\sigma$ -algebra of  $\mathcal{S}$ ), the  $U$ -statistic of kernels  $(h_{\mathbf{i}})$  is defined as

$$U_n((h_{\mathbf{i}})_{\mathbf{i} \in \text{Inc}^m}) = \sum_{\mathbf{i} \in \text{Inc}_n^m} h_{\mathbf{i}}(\xi_{\mathbf{i}}), \quad (1.1)$$

where  $\text{Inc}^m = \{(i_k)_{k=1}^m \mid 1 \leq i_1 < \dots < i_m\}$ ,  $\text{Inc}_n^m = \{\mathbf{i} = (i_k)_{k=1}^m \in \mathbb{N}^m, 1 \leq i_1 < i_2 < \dots < i_m \leq n\}$  and  $\xi_{\mathbf{i}} = (\xi_{i_1}, \dots, \xi_{i_m})$ .

When the kernels  $h_{\mathbf{i}}$  are index free, that is,  $h_{\mathbf{i}} = h$  and  $h(\xi_{\mathbf{i}})$  is integrable, it is known that  $U_n(h) / \binom{n}{2}$  converges almost surely to  $\mathbb{E}[h(\xi_{[1,m]})]$  (see [Hoeffding \(1948\)](#)), where  $\xi_{[1,m]} = (\xi_1, \dots, \xi_m)$  hence  $U_n(h) / \binom{n}{2}$  is an unbiased estimator of  $\mathbb{E}[h(\xi_{[1,m]})]$ . Allowing the kernel  $h$  to vary with the index opens a wider range of applications, including weighted  $U$ -statistics and also incomplete  $U$ -statistics.

In this paper, we will consider  $U$ -statistics taking values in a separable Hilbert space that will be denoted by  $\mathbb{H}$ . We provide an inequality for the tail of  $\|U_n(h_i)\|_{\mathbb{H}}$ . The upper bound is a addition of an exponential term with an other one which is expressed in terms via a tail of a sum of squared norms of some random variables linked with the kernels  $h_i$  and the sequence  $(\xi_i)_{i \geq 1}$ .

Exponential inequalities for  $U$ -statistics of i.i.d. data have been obtained by [Arcones \(1995\)](#), [Bickel and Ritov \(1995\)](#), [Borisov \(1990\)](#), [Christofides \(1991\)](#), [Giné et al. \(2000\)](#), [Houdré and Reynaud-Bouret \(2003\)](#) for bounded kernels, [Wang \(1998\)](#) for kernels having finite exponential moments.

For dependent sequences, the following results are available for bounded kernels. In the mixing case, a deviation inequality has been established by [Han \(2018\)](#) for fixed kernel under an assumption on  $\beta$ -mixing coefficients, and conditional  $\alpha$ -mixing coefficients. [Borisov and Volodko \(2009\)](#); [Borisov and Bystrov \(2019\)](#) dealt with the  $\phi$ -mixing case. The case of  $\alpha$ -mixing data combined with a control on the decay of the Fourier transform of the kernel was treated by [Shen et al. \(2020\)](#). [Duchemin et al. \(2023\)](#) gave exponential inequalities for  $U$ -statistics of order two, of varying kernel case where data is a function of a Markov chain.

One of the main motivations of studying Hilbert space valued  $U$ -statistic is to treat spatial sign for robust tests (see [Chakraborty and Chaudhuri \(2017\)](#); [Wegner and Wendler \(2023\)](#); [Jiang et al. \(2023\)](#)), where the sign function is replaced by the map sending the origin to itself and  $u \mapsto u/\|u\|_{\mathbb{H}}$  if  $u \neq 0$ . Wilcoxon-Mann-Whitney-type test are also studied in [Chakraborty and Chaudhuri \(2015\)](#). The consideration of  $U$ -statistics with varying kernel allows in particular to treat weighted  $U$ -statistics. These ones are widely used for change-point tests by [Dehling et al. \(2022\)](#); [Hu and Wang \(2023\)](#), ranking problems by [Cléménçon et al. \(2016, 2008\)](#). Gini mean-differences, that is, when  $m = 2$ ,  $S = \mathbb{H}$  and  $h_{i_1, i_2}(u_1, u_2) = \|u_1 - u_2\|_{\mathbb{H}}$ , when the sequence  $(\xi_i)_{i \geq 1}$  takes values in a separable Hilbert space can be useful in the study of functional data. The latter has also been done via  $U$ -statistics in [Hu et al. \(2023\)](#). Empirical  $U$ -statistics have also been studied (see [Ruymgaart and van Zuijlen \(1992\)](#) for the i.i.d. case, [Ragbi \(1998, 1997\)](#); [Harel et al. \(1994\)](#) in the mixing case). They can be viewed as random elements of an  $\mathbb{L}^2$  space hence the deviation inequalities we established can be used.

Our results deal with Hilbert space valued  $U$ -statistics, which are not the most general Banach spaces where  $U$ -statistics are involved (see for instance [Adamczak \(2006\)](#) and [Giraudo \(2024b\)](#), where type 2, respectively, smooth Banach spaces are considered). Let us explain why our work does not address the latter case. Our argument rests on a deviation inequality for Hilbert valued martingales. Using the dimension reduction explained in [Ding et al. \(2013\)](#), section 4, such an inequality could be established for smooth Banach spaces (like in [Luo \(2022\)](#)), at the cost of a constant that depends on  $\mathbb{B}$  through a constant  $C$  such that for each  $x, y \in \mathbb{B}$ ,

$$\|x + y\|_{\mathbb{B}}^p \leq \|x\|_{\mathbb{B}}^p + pJ_x(y) + C\|y\|_{\mathbb{B}}^p, \quad (1.2)$$

where  $1 < p \leq 2$  and  $J_x: \mathbb{B} \rightarrow \mathbb{R}$  is linear and continuous. We would like to use an induction argument and a change of Banach space, that is, taking  $\mathbb{B}^N$  like in [Giraudo \(2024a\)](#) but the constant  $C$  in (1.2) may depend on  $N$  (this was not an issue in the aforementioned paper, since the involved constant is related to a martingale moment inequality and is the same for  $\mathbb{B}$  and  $\mathbb{B}^N$ ).

## 1.2 An exponential inequality for degenerate $U$ -statistics

Our approach to derive deviation inequalities rests on martingale methods. In general,  $U$ -statistic of the form (1.1) are not sums of martingale differences in some sense. In order to use such properties for each index of summation, we will need the following assumption.

**Definition 1.1.** *The  $U$ -statistic  $U_n((h_i))$  defined as in (1.1) is degenerate if for each  $\mathbf{i} \in \text{Inc}^m$  and  $j_0 \in \llbracket 1, m \rrbracket$ ,*

$$\mathbb{E} [h_{\mathbf{i}}(\xi_{\llbracket 1, m \rrbracket}) \mid \xi_j, j \in \llbracket 1, m \rrbracket \setminus \{j_0\}] = 0 \text{ a.s.}, \quad (1.3)$$

where  $\llbracket a, b \rrbracket = \{k \in \mathbb{N}, a \leq k \leq b\}$  and  $\xi_{\llbracket 1, m \rrbracket} = (\xi_1, \dots, \xi_m)$ .

We will see later what can be done if (1.3) does not hold. The inequality for degenerate  $U$ -statistics reads as follows.

**Theorem 1.2.** *Let  $m \geq 1$ . There exist constants  $A_m$ ,  $B_m$  and  $C_m$  such that if  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  is a separable Hilbert space,  $(\xi_i)_{i \geq 1}$  is an i.i.d. sequence taking values in a measurable space  $(S, \mathcal{S})$ ,  $h_i: S^m \rightarrow \mathbb{H}$  are measurable functions for which (1.3) holds, then*

$$\begin{aligned} \mathbb{P} \left( \max_{m \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^m} h_{\mathbf{i}}(\xi_i) \right\|_{\mathbb{H}} > x \right) &\leq A_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m}} \right) \\ &+ B_m \int_1^\infty u (1 + \log u)^{\frac{m(m+1)}{2}-1} \mathbb{P} \left( \sqrt{\sum_{\mathbf{i} \in \text{Inc}_N^m} \|h_{\mathbf{i}}(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2} > C_m y u \right) du, \end{aligned} \quad (1.4)$$

where

$$\xi_{\mathbf{i}}^{\text{dec}} = (\xi_{i_1}^{(1)}, \dots, \xi_{i_m}^{(m)})$$

and  $(\xi_i^{(\ell)})_{i \geq 1}$ ,  $\ell \in \llbracket 1, m \rrbracket$ , are independent copies of  $(\xi_i)_{i \geq 1}$ .

Let us make some comments on Theorem 1.2.

When  $M_{\mathbf{i}} := \sup_{s_1, \dots, s_m \in S} \|h_{\mathbf{i}}(s_1, \dots, s_m)\|_{\mathbb{H}}$  is finite, the right hand side of (1.4) vanishes with the choice  $y = \sqrt{\sum_{\mathbf{i} \in \text{Inc}_N^m} M_{\mathbf{i}}^2} / C_m$  hence we get

$$\mathbb{P} \left( \max_{m \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^m} h_{\mathbf{i}}(\xi_i) \right\|_{\mathbb{H}} > x \right) \leq A_m \exp \left( -x^{\frac{2}{m}} C_m^{\frac{2}{m}} \left( \sum_{\mathbf{i} \in \text{Inc}_N^m} M_{\mathbf{i}}^2 \right)^{-m} \right). \quad (1.5)$$

It is also known by Arcones (1995) that when  $h = h_{\mathbf{i}}$  bounded and degenerate with respect to  $(\xi_i)_{i \geq 1}$ , the exponent  $2/m$  in the exponential term is optimal.

Notice also that the right hand side of (1.4) is finite if and only if for each  $\mathbf{i} \in \text{Inc}_n^m$ ,

$$\mathbb{E} \left[ \left\| h_{\mathbf{i}}(\xi_{\llbracket 1, m \rrbracket}) \right\|_{\mathbb{H}}^2 (1 + \log (\|h_{\mathbf{i}}(\xi_{\llbracket 1, m \rrbracket})\|_{\mathbb{H}}))^{\frac{m(m+1)}{2}-1} \right] < \infty.$$

In particular, this requires a bit more than the existence of a finite second moment. This is not a restriction at all, since in the application cases, the random variable  $\|h_{\mathbf{i}}(\xi_{[1,m]})\|_{\mathbb{H}}$  is supposed to admit finite exponential moments. When we only have finite moment of order between 1 and 2, one can use the deviation inequality presented in [Giraud \(2024b\)](#).

Let us explain how Theorem 1.2 can be applied. The value of  $x$  will be imposed by the application we have in mind. Moreover, for a fixed  $x$ , the right hand side of (1.4) is the sum of two terms, an exponential one, which is increasing in  $y$  and a tail one, which is non-increasing in  $y$ . Usually, we will choose  $y$  depending on  $N$  and  $x$  such that  $\exp\left(-(x/y)^{2/m}\right)$  has the wanted decay, which will furnish a condition on the tail of  $\sum_{\mathbf{i} \in \text{Inc}_N^m} \|h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}})\|_{\mathbb{H}}^2$ . However, the tail of such random variable is not easy to control. We will provide some examples where it will be possible to bound it.

## 2 Applications

The result of Theorem 1.2 apply when the kernels are degenerate. In the next examples, we will see which kind of bound can be obtained for non-necessarily degenerate kernels, and also with a bound involving the random variables  $\xi_{[1,m]}$  instead of  $\xi_i$ ,  $1 \leq i \leq N$ .

### 2.1 Index free kernel

In this Subsection, we consider the case where the functions  $h_{\mathbf{i}}$  do not depend on  $\mathbf{i}$  and will be simply denoted as  $h$ . The corresponding  $U$ -statistic will be expressed as

$$U_{m,n}(h) := \sum_{\mathbf{i} \in \text{Inc}_n^m} h(\xi_{\mathbf{i}}).$$

The degeneracy assumption (1.3) reads as for each  $j_0 \in [1, m]$ ,

$$\mathbb{E} [h(\xi_{[1,m]}) \mid \xi_{[1,m] \setminus \{j_0\}}] = 0 \text{ a.s.}$$

Of course, such an assumption has no reason to take place in general. We thus need to decompose  $h$  into a sum of functions of less than  $m$ -variables for which the corresponding  $U$ -statistic is degenerate.

**Definition 2.1.** Let  $(S, \mathcal{S})$  be a measurable space and let  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  be a separable Hilbert space. We say that the kernel  $h: S^m \rightarrow \mathbb{H}$  is symmetric if for each  $s_1, \dots, s_m \in \mathbb{H}$  and each bijective map  $\sigma: [1, m] \rightarrow [1, m]$ ,  $h(s_{\sigma(1)}, \dots, s_{\sigma(m)}) = h(s_1, \dots, s_m)$ .

Define for  $k \in [1, m]$  the kernels

$$h_k: S^k \rightarrow \mathbb{H}, h_k(s_1, \dots, s_k) = \sum_{j=0}^k (-1)^{k-j} \sum_{(u_\ell)_{\ell=1}^j \in \text{Inc}_k^j} \mathbb{E} [h(s_{u_1}, \dots, s_{u_j}, \xi_{j+1}, \dots, \xi_m)] \quad (2.1)$$

and  $h_0 = \mathbb{E}[h(\xi_1, \dots, \xi_m)]$ . In this way, the following equality, known as the Hoeffding's decomposition (see [Hoeffding \(1948\)](#)), takes place:

$$U_{m,n}(h) = \sum_{k=0}^m \binom{m}{k} \frac{\binom{n}{m}}{\binom{n}{k}} U_{k,n}(h_k). \quad (2.2)$$

When  $h$  is degenerate, the terms of index  $k \in \llbracket 0, m-1 \rrbracket$  vanish. Also, observe that for each  $k$ ,  $U_{k,n}(h_k)$  is a degenerate  $U$ -statistic and that if  $\mathbb{E}[h(\xi_{\llbracket 1, m \rrbracket} \mid \xi_{\llbracket 1, k \rrbracket})] = 0$ , then  $h_k(\xi_{\llbracket 1, k \rrbracket}) = 0$  almost surely.

**Definition 2.2.** Let  $m \geq 1$ , let  $(S, \mathcal{S})$  be a measurable space, let  $(\xi_i)_{i \geq 1}$  be an i.i.d. sequence taking values in  $S$  and let  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  be a separable Hilbert space. We say that the kernel  $h: S^m \rightarrow \mathbb{H}$  is degenerate of order  $d$  if for each  $k \in \llbracket 0, d-1 \rrbracket$ ,  $U_{k,n}(h_k) = 0$ , where  $h_k$  is defined as in (2.1).

It is clear that in view of (2.2) that if  $h$  is degenerate of order  $d$ , then

$$U_{m,n}(h) = \sum_{k=d}^m \binom{m}{k} \frac{\binom{n}{m}}{\binom{n}{k}} U_{k,n}(h_k). \quad (2.3)$$

**Theorem 2.3.** Let  $m \geq 1$ . There exists constants  $A_m$ ,  $B_m$  and  $C_m$  such that if  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  is a separable Hilbert space,  $(\xi_i)_{i \geq 1}$  is an i.i.d. sequence taking values in a measurable space  $(S, \mathcal{S})$ ,  $h: S^m \rightarrow \mathbb{H}$  is symmetric and degenerate of order  $d$ , then for each positive  $x$  and  $y$ ,

$$\begin{aligned} & \mathbb{P} \left( \max_{m \leq n \leq N} \|U_{m,n}(h)\|_{\mathbb{H}} > x N^{m-\frac{d}{2}} \right) \\ & \leq A_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{d}} \right) + B_m \sum_{k=d}^m \int_1^\infty \mathbb{P} \left( H_k > C_m u N^{\frac{k-d}{2}} x^{1-\frac{k}{d}} y^{\frac{k}{d}} \right) u (1 + \log u)^{\frac{m(m+1)}{2}} du, \end{aligned} \quad (2.4)$$

where

$$H_k := \mathbb{E} \left[ \|h(\xi_{\llbracket 1, m \rrbracket})\|_{\mathbb{H}} \mid \xi_{\llbracket 1, k \rrbracket} \right].$$

This result relates to Corollary 1.1 in [Giraud \(2021\)](#) in the following way. Our result deals with Hilbert-valued  $U$ -statistics, while that in [Giraud \(2021\)](#) is restricted to the real-valued case. However, in this latter case, the constants are explicit which is not the case in (2.4).

Rates of convergence in the law of large numbers can be derived in exactly the same way as it was done in [Giraud \(2021\)](#), replacing the absolute value by the Hilbert norm. We will therefore simply state them, without proof.

**Corollary 2.4.** Let  $(S, \mathcal{S})$  be a measurable space,  $m \geq 2$ ,  $h: S^m \rightarrow \mathbb{R}$  be a symmetric function, and let  $(\xi_i)_{i \geq 1}$  be an i.i.d. sequence of random variables with values in  $S$ . Suppose that  $h$  is degenerate of order  $d$ . If there exists a positive  $\gamma > 0$  such that  $M := \sup_{t > 0} \exp(t^\gamma) \mathbb{P}(\|h(\xi_{\llbracket 1, m \rrbracket})\|_{\mathbb{H}} > t)$  is finite, then for all  $x > 0$ ,

$$\mathbb{P} \left( \max_{m \leq n \leq N} \|U_{m,n}(h)\|_{\mathbb{H}} > N^m x \right) \leq K_1 \exp \left( -K_2 N^{\frac{d\gamma}{2+d\gamma}} x^{\frac{2\gamma}{2+d\gamma}} \right),$$

where  $K_1$  and  $K_2$  depend on  $\gamma$ ,  $M$  and  $m$ .

## 2.2 Weighted sums

In this Subsection, we provide results for weighted sums of non-necessarily degenerate  $U$ -statistics.

When  $h_{\mathbf{i}}$  does not depend on  $\mathbf{i}$  or more generally when  $h_{\mathbf{i}} = T_{\mathbf{i}}(h)$ , where  $T_{\mathbf{i}}: \mathbb{H} \rightarrow \mathbb{H}$  is a bounded linear operator and  $h$  is degenerate, it is possible to control the tail of  $\sum_{\mathbf{i} \in \text{Inc}_N^m} \|h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}})\|_{\mathbb{H}}^2$  in terms of the operator norms of  $T_{\mathbf{i}}$ , defined as  $\|T_{\mathbf{i}}\|_{\mathcal{B}(\mathbb{H})} = \sup \{\|T_{\mathbf{i}}(h)\|_{\mathbb{H}}, h \in \mathbb{H}, \|h\|_{\mathbb{H}} = 1\}$ , and the tail of  $\|h(\xi_{[1,m]})\|_{\mathbb{H}}$ .

When  $h$  is not necessarily degenerate, one can still do the Hoeffding's decomposition and get that

$$\sum_{\mathbf{i} \in \text{Inc}_n^m} T_{\mathbf{i}}(h(\xi_{\mathbf{i}})) = \sum_{k=d}^m \sum_{\mathbf{i} \in \text{Inc}_n^k} a_{\mathbf{i}}^{(n,k)}(h_k(\xi_{\mathbf{i}})), \quad (2.5)$$

where  $h_k$  is defined as in (2.1) and for  $\mathbf{i} = (i_1, \dots, i_k) \in \text{Inc}_n^k$ ,

$$a_{\mathbf{i}}^{(n,k)} = \sum_{\mathbf{j}=(j_1, \dots, j_m) \in \text{Inc}_n^m, \{i_1, \dots, i_k\} \subset \{j_1, \dots, j_m\}} T_{\mathbf{j}}.$$

When  $m = 2$ , one has

$$a_{\mathbf{i}}^{(n,1)} = \sum_{j=1}^n (T_{i,j} + T_{j,i}),$$

with the convention that  $T_{j,i} = 0$  if  $j \leq i$ . For  $m = 3$ , one has

$$a_{i_1, i_2}^{(n,2)} = \sum_{j=1}^n (T_{i_1, i_2, j} + T_{i_1, j, i_2} + T_{j, i_1, i_2})$$

$$a_{i_1}^{(n,1)} = \sum_{1 \leq j_1 < j_2 \leq n} (T_{i_1, j_1, j_2} + T_{j_1, i_1, j_2} + T_{j_1, j_2, i_1}),$$

with the convention that  $T_{k_1, k_2, k_3} = 0$  if  $k_1 < k_2 < k_3$  does not hold.

Since the operators  $a_{\mathbf{i}}^{(n,k)}$  depend on  $n$ , we cannot formulate a deviation inequality for the running maximum. A deviation inequality for weighted and non-necessarily degenerate  $U$ -statistics reads as follows.

**Corollary 2.5.** *Let  $m \geq 1$ . There exist constants  $a_m$ ,  $b_m$  and  $c_m$  such that if  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  is a separable Hilbert space,  $(\xi_i)_{i \geq 1}$  is an i.i.d. sequence taking values in a measurable space  $(S, \mathcal{S})$ ,  $h: S^m \rightarrow \mathbb{H}$  is a measurable function which is degenerate of order  $d$  and  $T_{\mathbf{i}}: \mathbb{H} \rightarrow \mathbb{H}$  are bounded linear operators, then for each positive  $x$  and  $y$ ,*

$$\mathbb{P} \left( \left\| \sum_{\mathbf{i} \in \text{Inc}_n^m} T_{\mathbf{i}}(h(\xi_{\mathbf{i}})) \right\|_{\mathbb{H}} > x n^{\frac{m-d}{2}} \sqrt{\sum_{\mathbf{i} \in \text{Inc}_n^m} \|T_{\mathbf{i}}\|_{\mathcal{B}(\mathbb{H})}^2} \right) \leq a_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m}} \right) + b_m \int_1^\infty u (1 + \log u)^{\frac{m(m+1)}{2}} \mathbb{P}(H > c_m y u) du, \quad (2.6)$$

where

$$H = \|h(\xi_{[1,m]})\|_{\mathbb{H}}.$$

## 2.3 Incomplete $U$ -statistics

The computation of a  $U$ -statistic of order  $m$  based on a sample of size  $n$  requires a computation of a number of terms of order  $n^m$ , which can be high in practise. For this reason, [Blom \(1976\)](#) introduced the so-called incomplete  $U$ -statistics, whose rough idea is to put a random weight equal to 0 or 1, reducing the number of terms that one has to compute. Here are the most common ways of defining the random weights.

- Sampling without replacement: we pick without replacement  $N$   $m$ -uples which belong to  $\text{Inc}^m$ .
- Sampling with replacement: we pick with replacement  $N$   $m$ -uples which belong to  $\text{Inc}^m$ .
- Bernoulli sampling: consider for each  $\mathbf{i} \in \text{Inc}_n^m$ , a random variable  $Z_{n,\mathbf{i}}$  taking the value 1 with probability  $p_n$  and 0 with probability  $1 - p_n$ . Moreover, we assume that the family  $(Z_{n,\mathbf{i}})_{\mathbf{i} \in \text{Inc}_n^m}$  is independent and also independent of the sequence  $(\xi_i)_{i \in \mathbb{Z}}$ .

The weak convergence of incomplete  $U$ -statistics has been established in [Janson \(1984\)](#). Rates in the law of large numbers were obtained by [Nasari \(2012\)](#) and for  $\mathbb{L}^1$  convergence by [Dürre and Paindaveine \(2023\)](#). Recent papers include incomplete  $U$ -statistics in a high dimensional setting, like [Chen and Kato \(2019\)](#), and [Löwe and Terveer \(2021\)](#) obtained a central limit theorem when the data is based on a triangular array.

Our first result deals with sampling with or without replacement.

**Corollary 2.6** (Deviation inequality for sampling with and without replacement). *Let  $m \geq 1$ . There exist constants  $a_m$ ,  $b_m$  and  $c_m$  such that if  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  is a separable Hilbert space,  $(\xi_i)_{i \geq 1}$  is an i.i.d. sequence taking values in a measurable space  $(S, \mathcal{S})$ ,  $h: S^m \rightarrow \mathbb{H}$  is a measurable function which is degenerate of order  $d$ ,  $N \geq 1$  and  $n \geq m$  are integers and  $(Z_{n,\mathbf{i}})_{\mathbf{i} \in \text{Inc}_n^m}$  is a collection of random variables taking the values 0 or 1 which is independent of  $(\xi_i)_{i \geq 1}$  and such that  $\sum_{\mathbf{i} \in \text{Inc}_n^m} Z_{n,\mathbf{i}} = N$  and  $x, y > 0$ , then*

$$\mathbb{P} \left( \left\| \sum_{\mathbf{i} \in \text{Inc}_n^m} Z_{n,\mathbf{i}} h(\xi_{\mathbf{i}}) \right\|_{\mathbb{H}} > x \sqrt{N} \sqrt{\min \{N, n^{m-d}\}} \right) \leq a_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m}} \right) + b_m \int_1^\infty u (1 + \log u)^{\frac{m(m+1)}{2}} \mathbb{P}(H > c_m y u) du. \quad (2.7)$$

Observe that the sampling with or without replacement is packed gives the same result. The only crucial thing is the numbers of random weights that are not zero.

Also, observe that the normalization  $\sqrt{N} \sqrt{\min \{N, n^{m-d}\}}$  is decreasing with  $d$ : if the  $U$ -statistic is degenerate of order  $m$ , we only need a normalization by  $\sqrt{N}$  in order to get a bounded independent of  $N$  and  $n$ . In the other extreme case, that is, when  $d = 1$ , which corresponds to the case where  $h_{\mathbf{i}}(\xi_{\mathbf{i}})$  is centered, if one takes  $N \leq n^{m-1}$  (which is logical if one wants to reduce significantly the number of elements), then one needs the much stronger normalization  $N$ .

A deviation inequality for Bernoulli sampling reads as follows.



**Corollary 2.7** (Deviation inequality for Bernoulli sampling). *Let  $m \geq 1$ . There exist constants  $a_m, b_m$  and  $c_m$  such that if  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  is a separable Hilbert space,  $(\xi_i)_{i \geq 1}$  is an i.i.d. sequence taking values in a measurable space  $(S, \mathcal{S})$ ,  $h: S^m \rightarrow \mathbb{H}$  is a measurable function which is degenerate of order  $d$ ,  $N \geq 1$  and  $n \geq m$  are integers and  $(Z_{n,i})_{i \in \text{Inc}_n^m}$  is a collection of random variables taking the values 0 or 1 with respective probabilities  $1 - p_n$  and  $p_n$ , which is independent of  $(\xi_i)_{i \geq 1}$  and  $x, y > 0$ , then*

$$\begin{aligned} & \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^m} Z_{n,i} h(\xi_i) \right\|_{\mathbb{H}} > x n^m \sqrt{p_n} \sqrt{\min \{p_n, n^{-d}\}} \right) \\ & \leq a_m \exp \left( -\frac{n^m p_n^2}{2} \right) + a_m \exp \left( -\left( \frac{x}{y} \right)^{\frac{2}{m}} \right) + b_m \int_1^\infty u (1 + \log u)^{\frac{m(m+1)}{2}} \mathbb{P}(H > c_m y u) du. \end{aligned} \quad (2.8)$$

Let us compare Corollaries 2.6 with 2.7. The right hand side of (2.7) is similar, up to the constants, to the sum of the second and third terms of the right hand side of (2.8). These quantities are independent of  $n$  and of the involved parameters of the sampling methods. The required normalizations to have a right hand side independent of  $n, N$  or  $p_n$  (up to the first term for (2.8)) depend on  $d$  and are the weakest if  $d = m$  and the strongest if  $d = 1$ . When  $N = n^m p_n$ , the normalizing terms coincide. Nevertheless, an other exponential term appears in (2.8) due to the control of the probability that the sum of the weights is bigger than some number.

Let us point out that a Bernstein type exponential inequality for incomplete real valued  $U$ -statistics has been established by Maurer (2022). The results deals with sampling with replacement, bounded kernels and the author does not make any assumption on degeneracy. The results are expressed in terms of the number of  $m$ -uples among the chosen  $N$  that contain a specific value  $k$ , or that contain both  $k$  and  $\ell$ . Therefore, this does not make Corollary 2.6 directly comparable with Maurer's results.

If we assume that  $H$  has a weak exponential moment, namely that for some positive  $\kappa_0$  and  $\alpha > 0$ ,

$$\sup_{t>0} \exp(\kappa_0 t^\alpha) \mathbb{P}(H > t) < \infty, \quad (2.9)$$

the bounds (2.7) and (2.8) can be simplified as follows.

For the sampling with or without replacement, one gets the existence of a constant  $\kappa > 0$  such that

$$\sup_{n \geq m} \sup_{x>0} \exp \left( \kappa x^{\frac{2\gamma}{2+m\gamma}} \right) \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^m} Z_{n,i} h(\xi_i) \right\|_{\mathbb{H}} > x \sqrt{N} \sqrt{\min \{N, n^{m-d}\}} \right) < \infty. \quad (2.10)$$

For Bernoulli sampling, there exists a constant  $K$  such that for each  $n, p_n \in (0, 1)$  and  $x > 0$ ,

$$\exp \left( - \min \left\{ \frac{n^m p_n}{2}, \kappa x^{\frac{2\gamma}{2+m\gamma}} \right\} \right) \mathbb{P} \left( \left\| \sum_{\mathbf{i} \in \text{Inc}_n^m} Z_{n,\mathbf{i}} h(\xi_{\mathbf{i}}) \right\|_{\mathbb{H}} > x n^m \sqrt{p_n} \sqrt{\min \{p_n, n^{-d}\}} \right) \leq K. \quad (2.11)$$

### 3 Proofs

#### 3.1 Proof of Theorem 1.2

Let us first explain the global idea of proof. It will be show that it is sufficient to prove Theorem 1.2 for decoupled  $U$ -statistics and only in the case  $x/y > 3^{m/2}$ . Then we proceed by induction over  $m$ . The induction step is performed first via an application of the martingale inequality given in Proposition A.5 for an appropriated choice of the filtration. Then, in order to use the induction assumption, we introduce a new Hilbert space so that we can apply the result for  $U$ -statistics of lower order.

In order to ease the notation of the exponent of the log appearing in the involved terms in the proof, we define

$$p_m = \frac{m(m+1)}{2} - 1, \quad m \geq 1.$$

In view of Corollary B.2, it suffices to prove Theorem 1.2 for decoupled  $U$ -statistics. Indeed, assume that for each  $m \geq 1$ , we can find constants  $A'_m$ ,  $B'_m$  and  $C'_m$  for which the inequality

$$\begin{aligned} \mathbb{P} \left( \max_{m \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^m} h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\mathbb{H}} > x \right) &\leq A'_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m}} \right) \\ &+ B'_m \int_1^\infty u (1 + \log u)^{p_m} \mathbb{P} \left( \sqrt{\sum_{\mathbf{i} \in \text{Inc}_N^m} \|h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}})\|_{\mathbb{H}}^2} > C'_m y u \right) du \end{aligned}$$

holds for each Hilbert space  $\mathbb{H}$  and each functions  $h_{\mathbf{i}}$  which are degenerate with respect to  $(\xi_i)_{i \geq 1}$ . Applying Corollary B.2, we derive that

$$\begin{aligned} \mathbb{P} \left( \max_{m \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^m} h_{\mathbf{i}}(\xi_{\mathbf{i}}) \right\|_{\mathbb{H}} > x \right) &\leq A'_m K_m \exp \left( - \left( \frac{x}{y K_m} \right)^{\frac{2}{m}} \right) \\ &+ B'_m K_m \int_1^\infty u (1 + \log u)^{p_m} \mathbb{P} \left( \sqrt{\sum_{\mathbf{i} \in \text{Inc}_N^m} \|h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}})\|_{\mathbb{H}}^2} > C'_m y u \right) du \end{aligned}$$

and replacing  $y$  by  $y/K_m$ , we can see that letting  $A_m := A'_m K_m$ ,  $B_m = B'_m K_m$  and  $C_m = C'_m/K_m$ , (1.4) holds.

We can also reduce the proof to the case  $x/y > 3^{m/2}$ , by replacing  $A_m$  by  $\max\{A_m, e^3\}$ , so that the contribution of  $\max\{A_m, e^3\} \exp\left(-(x/y)^{2/m}\right)$  makes the inequality trivial for the range of  $(x, y)$  for which  $x/y \leq 3^{m/2}$ .

Therefore, we will prove by induction on  $m \geq 1$  the assertion  $P(m)$  defined as follows: "there exists constants  $a_m, b_m$  and  $c_m$  such that for each measurable space  $(S, \mathcal{S})$ , each independent copies  $(\xi_i^k)$ ,  $k \in \llbracket 1, m \rrbracket$  of an i.i.d.  $S$ -valued sequence  $(\xi_i)_{i \geq 1}$ , each separable Hilbert space  $(\mathbb{H}, \langle \cdot, \cdot \rangle)_{\mathbb{H}}$ , each collection  $h_{\mathbf{i}}: S^m \rightarrow \mathbb{H}$ ,  $\mathbf{i} \in \text{Inc}^m$  of measurable maps such that for each  $\mathbf{i} \in \text{Inc}^m$  and each  $j_0 \in \llbracket 1, m \rrbracket$ ,

$$\mathbb{E} [h_{\mathbf{i}}(\xi_{\llbracket 1, m \rrbracket}) \mid \sigma(\xi_j, j \in \llbracket 1, m \rrbracket \setminus \{j_0\})] = 0, \quad (3.1)$$

the following inequality holds for each  $x, y > 0$  satisfying  $x/y > 3^{m/2}$  and each  $N \geq m$ :

$$\begin{aligned} \mathbb{P} \left( \max_{m \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^m} h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\mathbb{H}} > x \right) &\leq a_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m}} \right) \\ &+ b_m \int_1^\infty u (1 + \log u)^{p_m} \mathbb{P} \left( \sqrt{\sum_{\mathbf{i} \in \text{Inc}_N^m} \|h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}})\|_{\mathbb{H}}^2} > c_m y u \right) du. \end{aligned}$$

For  $m = 1$ , this follows directly from Proposition A.3, since (3.1) implies that the considered  $U$ -statistic is a sum of independent centered random variables.

Assume that  $P(m)$  holds for some  $m \geq 1$  and let us show  $P(m+1)$ . Let  $(S, \mathcal{S})$  be a measurable space,  $(\xi_i)_{i \geq 1}$  an i.i.d. sequence of  $S$ -valued random variables, let  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  be a Hilbert space and let  $h_{\mathbf{i}, i_{m+1}}: S^{m+1} \rightarrow \mathbb{H}$  be measurable functions such that for each  $j_0 \in \llbracket 1, m+1 \rrbracket$ ,

$$\mathbb{E} [h_{\mathbf{i}, i_{m+1}}(\xi_{\llbracket 1, m+1 \rrbracket}) \mid \sigma(\xi_j, j \in \llbracket 1, m+1 \rrbracket \setminus \{j_0\})] = 0 \text{ a.s.} \quad (3.2)$$

Let  $N \geq m+1$  and  $x, y > 0$  such that  $x/y > 3^{(m+1)/2}$ . Let us show that there are constants  $a_{m+1}, b_{m+1}$  and  $c_{m+1}$  depending only on  $m+1$  such that

$$\begin{aligned} \mathbb{P} \left( \max_{m+1 \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^{m+1}} h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\mathbb{H}} > x \right) &\leq a_{m+1} \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m+1}} \right) \\ &+ b_{m+1} \int_1^\infty u (1 + \log u)^{p_{m+1}} \mathbb{P} \left( \sqrt{\sum_{\mathbf{i} \in \text{Inc}_N^{m+1}} \|h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}})\|_{\mathbb{H}}^2} > c_{m+1} y u \right) du. \end{aligned} \quad (3.3)$$

Define for  $j \geq m+1$  the random variable

$$D_j := \sum_{\mathbf{i} \in \text{Inc}_{j-1}^m} h_{\mathbf{i}, j}(\xi_{\mathbf{i}}^{\text{dec}}, \xi_j^{(m+1)})$$

and the  $\sigma$ -algebras

$$\begin{aligned} \mathcal{F}_m &:= \sigma(\xi_{\mathbf{i}}^{\text{dec}}, \mathbf{i} \in \text{Inc}_{N-1}^m), \\ \mathcal{F}_j &= \mathcal{F}_m \vee \sigma(\xi_u^{(m+1)}, u \leq j), j \geq m+1. \end{aligned}$$

Recall the following elementary property of conditional expectations: if  $Y$  is an integrable random variable on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and  $\mathcal{G}, \mathcal{H}$  are two sub- $\sigma$ -algebras of  $\mathcal{F}$  such that  $\mathcal{H}$  is independent of  $\sigma(Y) \vee \mathcal{G}$ , then

$$\mathbb{E}[Y \mid \mathcal{G} \vee \mathcal{H}] = \mathbb{E}[Y \mid \mathcal{G}]. \quad (3.4)$$

Let  $j \geq m+1$  be fixed. Then the following equality holds:

$$\mathbb{E}[D_j \mid \mathcal{F}_{j-1}] = \sum_{\mathbf{i} \in \text{Inc}_{j-1}^m} \mathbb{E}\left[h_{\mathbf{i},j}\left(\xi_{\mathbf{i}}^{\text{dec}}, \xi_j^{(m+1)}\right) \mid \mathcal{F}_{j-1}\right].$$

Using equality (3.4) for a fixed  $\mathbf{i} \in \text{Inc}_{j-1}^m$  with  $Y = h_{\mathbf{i},j}\left(\xi_{\mathbf{i}}^{\text{dec}}, \xi_j^{(m+1)}\right)$ ,  $\mathcal{G} = \sigma(\xi_{\mathbf{i}}^{\text{dec}})$  and

$$\mathcal{H} := \sigma(\xi_{\mathbf{i}'}^{\text{dec}}, \mathbf{i}' \in \text{Inc}_{N-1}^m \setminus \{\mathbf{i}\}) \vee \sigma(\xi_u^{(m+1)}, u \leq j-1),$$

furnishes the equality

$$\mathbb{E}[D_j \mid \mathcal{F}_{j-1}] = \sum_{\mathbf{i} \in \text{Inc}_{j-1}^m} \mathbb{E}\left[h_{\mathbf{i},j}\left(\xi_{\mathbf{i}}^{\text{dec}}, \xi_j^{(m+1)}\right) \mid \sigma(\xi_{\mathbf{i}}^{\text{dec}})\right],$$

which equals 0 by (3.2) hence  $(D_j)_{j \geq m+1}$  is a martingale difference sequence with respect to the filtration  $(\mathcal{F}_j)_{j \geq m}$ . Moreover, observe that

$$\mathbb{E}[\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_{j-1}] = \mathbb{E}[\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_m \vee \sigma(\xi_u^{(m+1)}, u \leq j-1)]$$

hence an other application of (3.4) to  $\mathcal{G} = \mathcal{F}_m$  and  $\mathcal{H} = \sigma(\xi_u^{(m+1)}, u \leq j-1)$  gives  $\mathbb{E}[\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_{j-1}] = \mathbb{E}[\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_m]$ . Using Proposition A.5 with  $x_1 = x$  and  $y_1 = y^{\frac{1}{m+1}} x^{\frac{m}{m+1}}$  gives

$$\begin{aligned} \mathbb{P}\left(\max_{m+1 \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^{m+1}} h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\mathbb{H}} > x\right) &= \mathbb{P}\left(\max_{m+1 \leq n \leq N} \left\| \sum_{j=m+1}^n D_j \right\|_{\mathbb{H}} > x\right) \\ &\leq 4 \exp\left(-\left(\frac{x}{y}\right)^{\frac{2}{m+1}}\right) + 4 \int_1^\infty u \mathbb{P}\left(\sqrt{\sum_{j=m+1}^N \|D_j\|_{\mathbb{H}}^2} > y^{\frac{1}{m+1}} x^{\frac{m}{m+1}} \frac{u}{8}\right) du, \end{aligned} \quad (3.5)$$

which can be rewritten, by independence between  $(\xi_{\mathbf{i}}^{\text{dec}})_{\mathbf{i} \in \text{Inc}^m}$  and  $(\xi_j^{(m+1)})_{j=m+1}^N$ , as

$$\begin{aligned} \mathbb{P}\left(\max_{m+1 \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^{m+1}} h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\mathbb{H}} > x\right) &\leq 4 \exp\left(-\left(\frac{x}{y}\right)^{\frac{2}{m+1}}\right) \\ &+ 4 \int_1^\infty u \int_{S^{N-m}} \mathbb{P}\left(\sum_{j=m+1}^N \left\| \sum_{\mathbf{i} \in \text{Inc}_{j-1}^m} h_{\mathbf{i},j}(\xi_{\mathbf{i}}^{\text{dec}}, s_j) \right\|_{\mathbb{H}}^2 > y^{\frac{2}{m+1}} x^{\frac{2m}{m+1}} \frac{u^2}{64}\right) d\mathbb{P}_{\xi_{[m+1,N]}^{(m+1)}}(s_{[m+1,N]}) du, \end{aligned} \quad (3.6)$$

where for a function  $g: S^{N-m} \rightarrow \mathbb{R}$ ,

$$\int g(s_{m+1}, \dots, s_N) d\mathbb{P}_{\xi_{[m+1, N]}^{(m+1)}}(s_{[m+1, N]}) = \int g(s_{m+1}, \dots, s_N) d\mathbb{P}_{\xi_{m+1}}(s_{m+1}) \dots d\mathbb{P}_{\xi_N}(s_N).$$

In order to control the tail involved in the right hand side of the previous equation, we introduce the separable Hilbert space  $(\widetilde{\mathbb{H}}, \langle \cdot, \cdot \rangle_{\widetilde{\mathbb{H}}})$  which is defined as  $\widetilde{\mathbb{H}} = \left\{ (x_j)_{j \in [m+1, N]}, x_j \in \mathbb{H} \right\}$  and

$$\left\langle (x_j)_{j \in [m+1, N]}, (y_j)_{j \in [m+1, N]} \right\rangle_{\widetilde{\mathbb{H}}} = \sum_{j=m+1}^N \langle x_j, y_j \rangle_{\mathbb{H}},$$

and for fixed  $s_j, j \in [m+1, N]$ , we define the functions  $\widetilde{h_{\mathbf{i};(s_j)_{j \in [1, m]}}}: S^m \rightarrow \widetilde{\mathbb{H}}$  by

$$\widetilde{h_{\mathbf{i};(s_j)_{j \in [1, m]}}}(s_1, \dots, s_m) = \left( h_{\mathbf{i}, j}(s_1, \dots, s_m, s_j) \mathbf{1}_{\mathbf{i} \in \text{Inc}_{j-1}^m} \right)_{j \in [m+1, N]}. \quad (3.7)$$

Consequently, (3.6) can be rephrased as

$$\begin{aligned} \mathbb{P} \left( \max_{m+1 \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^{m+1}} h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\mathbb{H}} > x \right) &\leq 4 \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m+1}} \right) \\ &+ 4 \int_1^\infty u \int_{S^{N-m}} \mathbb{P} \left( \left\| \sum_{\mathbf{i} \in \text{Inc}_{N-1}^m} \widetilde{h_{\mathbf{i};(s_j)_{j \in [1, m]}}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\widetilde{\mathbb{H}}} > y^{\frac{1}{m+1}} x^{\frac{m}{m+1}} \frac{u}{8} \right) d\mathbb{P}_{\xi_{[m+1, N]}^{(m+1)}}(s_{[m+1, N]}) du. \end{aligned} \quad (3.8)$$

By assumption (3.2), the following equality takes place for  $\mathbb{P}_{\xi_m^{(m+1)}} \otimes \dots \otimes \mathbb{P}_{\xi_N^{(m+1)}}$  almost every  $(s_j)_{j \in [m+1, N]}$ :

$$\mathbb{E} \left[ \widetilde{h_{\mathbf{i};(s_j)_{j \in [m+1, N]}}}(\xi_{[1, m]}) \mid \xi_{[1, m] \setminus \{j_0\}} \right] = 0.$$

Therefore, an application of  $P(m)$  for such  $s_j$ , fixed  $u \geq 1$  and  $x$  and  $y$  replaced respectively by

$$x_2 = y^{\frac{1}{m+1}} x^{\frac{m}{m+1}} \frac{u}{8}, \quad y_2 = \frac{yu}{8(1 + \log u)^{\frac{m}{2}}},$$

gives

$$\begin{aligned} \mathbb{P} \left( \max_{m+1 \leq n \leq N} \left\| \sum_{\mathbf{i} \in \text{Inc}_n^{m+1}} h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\mathbb{H}} > x \right) &\leq 4 \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m+1}} \right) \\ &+ 2a_m \int_1^\infty u \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m+1}} (1 + \log u) \right) du + b_m \int_{S^{N-m}} \int_1^\infty \int_1^\infty uv (1 + \log v)^{p_m} \\ &\mathbb{P} \left( \sqrt{\sum_{\mathbf{i} \in \text{Inc}_{N-1}^m} \left\| \widetilde{h_{\mathbf{i};(s_j)_{j \in [m+1, N]}}}(\xi_{\mathbf{i}}^{\text{dec}}) \right\|_{\widetilde{\mathbb{H}}}^2} > \frac{yuv c_m}{8(1 + \log u)^{\frac{m}{2}}} \right) dudvd\mathbb{P}_{\xi_{[m+1, N]}^{(m+1)}}(s_{[m+1, N]}). \end{aligned} \quad (3.9)$$

Since  $x/y > 3^{(m+1)/2}$ , we get

$$2a_m \int_1^\infty u \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m+1}} (1 + \log u) \right) du \leq 2a_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m+1}} \right).$$

Moreover, using the expression of  $\|\cdot\|_{\mathbb{H}}$  and (3.7) allows us to derive from (3.9) that

$$\begin{aligned} & \mathbb{P} \left( \max_{m+1 \leq n \leq N} \left\| \sum_{i \in \text{Inc}_n^{m+1}} h_i(\xi_i^{\text{dec}}) \right\|_{\mathbb{H}} > x \right) \leq (4 + 2a_m) \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m+1}} \right) \\ & \quad + b_m \int_{S^{N-m}} \int_1^\infty \int_1^\infty uv (1 + \log v)^{p_m} \\ & \quad \mathbb{P} \left( \sum_{i \in \text{Inc}_{N-1}^m} \sum_{j=i_m}^N \|h_{i,j}(\xi_i^{\text{dec}}, s_j)\|_{\mathbb{H}}^2 > \frac{yuv c_m}{8(1 + \log u)^{m/2}} \right) dudv d\mathbb{P}_{\xi_{[m+1, N]}^{(m+1)}}(s_{[m+1, N]}) \\ & = (4 + 2a_m) \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m+1}} \right) + b_m \int_1^\infty \int_1^\infty uv (1 + \log v)^{p_m} g \left( \frac{yuv c_m}{(1 + \log u)^{m/2}} \right) dudv, \quad (3.10) \end{aligned}$$

where

$$g(t) = \mathbb{P} \left( \sqrt{\sum_{i \in \text{Inc}_N^{m+1}} \|h_i(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2} > \frac{t}{8} \right).$$

Doing the substitution  $w = u(1 + \log u)^{-m/2}v$  for a fixed  $u \geq 1$ , it suffices to prove the existence of constants  $\kappa_m$  and  $\kappa'_m$  for which for each  $w \in \mathbb{R}$ ,

$$\begin{aligned} I(w) &:= \int_1^\infty \mathbf{1}_{u(1+\log u)^{-m/2} < w} \frac{1}{u} (1 + \log u)^m \left( 1 + \log \left( \frac{w(1 + \log u)^{m/2}}{u} \right) \right)^{p_m} du \\ &\leq \kappa_m \mathbf{1}_{w > \kappa'_m} (1 + \log w)^{p_{m+1}}. \quad (3.11) \end{aligned}$$

Let us show (3.11). From the observation that there exists a constant  $\kappa_{1,m}$  depending only on  $m$  such that for each  $u \geq 1$ ,  $\sqrt{u} \leq \kappa_{1,m} u(1 + \log u)^{-m/2}$ , we derive that  $I(w)$  vanishes if  $w < (\kappa_{1,m})^{-1}$ , hence

$$I(w) \leq \mathbf{1}_{w > (\kappa_{1,m})^{-1}} \int_1^{w^2(\kappa_{1,m})^2} \frac{1}{u} (1 + \log u)^m \left( 1 + \log \left( \frac{w(1 + \log u)^{m/2}}{u} \right) \right)^{p_m} du.$$

Moreover, using that  $(1 + \log u)^{m/2}/u$  can be bounded independently of  $u \geq 1$ , we get for some constant  $\kappa_{2,m}$  that

$$\begin{aligned} I(w) &\leq \kappa_{2,m} \mathbf{1}_{w > 1} \int_1^{w^2(\kappa_{1,m})^2} \frac{1}{u} (1 + \log u)^m (1 + \log w)^{p_m} du \\ &\quad + \kappa_{2,m} \mathbf{1}_{(\kappa_{1,m})^{-1} < w \leq 1} \int_1^{(\kappa_{1,m})^2} \frac{1}{u} (1 + \log u)^m \left( 1 + \log \left( \frac{(1 + \log u)^{m/2}}{u} \right) \right)^{p_m} du \quad (3.12) \end{aligned}$$

which gives (3.11). This ends the proof of Theorem 1.2.

### 3.2 Proof of Theorem 2.3

Starting from (2.3), one has

$$\max_{m \leq n \leq N} \|U_{m,n}(h)\|_{\mathbb{H}} \leq \sum_{k=d}^m \binom{m}{k} \max_{k \leq n \leq N} \|U_{k,n}(h_k)\|_{\mathbb{H}} \frac{\binom{m}{n}}{\binom{m}{k}} \leq \kappa_m \sum_{k=d}^m N^{m-k} \max_{k \leq n \leq N} \|U_{k,n}(h_k)\|_{\mathbb{H}},$$

where  $h_k$  is defined as in (2.1) and  $\kappa_m$  depends only on  $m$ , hence

$$\mathbb{P} \left( \max_{m \leq n \leq N} \|U_{m,n}(h)\|_{\mathbb{H}} > x N^{m-\frac{d}{2}} \right) \leq \sum_{k=d}^m \mathbb{P} \left( \max_{k \leq n \leq N} \|U_{k,n}(h_k)\|_{\mathbb{H}} > x N^{k-\frac{d}{2}} \kappa_m / m \right). \quad (3.13)$$

For each  $k \in \llbracket d, m \rrbracket$ , we apply Theorem 1.2 in the following context:  $m_k = k$ ,  $h_i = h_k$ ,  $x_k = x N^{k-d/2} \kappa_m / m$  and  $y_k = N^{k-d/2} x^{1-k/d} \kappa_m / m$ . We obtain in view of (3.13) that

$$\begin{aligned} \mathbb{P} \left( \max_{m \leq n \leq N} \|U_{m,n}(h)\|_{\mathbb{H}} > x N^{m-\frac{d}{2}} \right) &\leq \sum_{k=d}^m A_k \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{d}} \right) \\ &\quad + B_k \int_1^\infty u (1 + \log u)^{\frac{m(m+1)}{2}-1} \mathbb{P} \left( \sum_{i \in \text{Inc}_N^k} \|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2 > \left( \frac{C_k}{m} N^{k-\frac{d}{2}} x^{1-\frac{k}{d}} \kappa_m u \right)^2 \right) du, \end{aligned}$$

where  $A_k, B_k, C_k, k \in \llbracket d+1, m \rrbracket$  are as in Theorem 1.2. Also, since the random variables  $\|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2$ ,  $i \in \text{Inc}_N^k$ , have the same distribution, one has

$$\sum_{i \in \text{Inc}_N^k} \|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2 \leq_{\text{conv}} N^k \|h_k(\xi_{[1,k]})\|_{\mathbb{H}}^2 \leq_{\text{conv}} N^k \kappa_k H_k, \quad (3.14)$$

where  $\leq_{\text{conv}}$  is defined as in Definition A.4 and  $\kappa_k$  depends only on  $k$ . Therefore, we infer from (A.8) that

$$\begin{aligned} \mathbb{P} \left( \max_{m \leq n \leq N} \|U_{m,n}(h)\|_{\mathbb{H}} > x N^{m-\frac{d}{2}} \right) &\leq \sum_{k=d}^m A_k \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{d}} \right) \\ &\quad + \sum_{k=d}^m B_k \int_1^\infty \int_1^\infty u (1 + \log u)^{p_m} \mathbb{P} \left( N^k \|h_k(\xi_{[1,k]})\|_{\mathbb{H}}^2 > \left( \frac{C_k}{m} N^{\frac{2k-d}{2}} x^{\frac{d-k}{d}} \kappa_m u \right)^2 \frac{v}{4} \right) dudv \\ &= \sum_{k=d}^m A_k \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{d}} \right) + \sum_{k=d}^m B_k \int_1^\infty \int_1^\infty u (1 + \log u)^{p_m} \mathbb{P} \left( H > \frac{C_k \kappa_m}{2m} N^{\frac{k-d}{2}} x^{\frac{d-k}{d}} u \sqrt{v} \right) dudv. \end{aligned}$$

Lemma 2.2 in Giraud (2021) gives

$$\int_1^\infty \int_1^\infty u (1 + \log u)^{p_m} g(u \sqrt{v}) dudv \leq \kappa_m \int_1^\infty u (1 + \log u)^{p_{m=1}} g(u / \kappa_m), \quad (3.15)$$

which concludes the proof of Theorem 2.3.

### 3.3 Proof of Corollary 2.5

Starting from the decomposition (2.5), one has

$$\mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^m} T_i(h(\xi_i)) \right\|_{\mathbb{H}} > x \right) \leq \sum_{k=d}^m \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_k^m} a_i^{(n,k)}(h_k(\xi_i)) \right\|_{\mathbb{H}} > x/m \right).$$

For each fixed  $k \in \llbracket d, m \rrbracket$ , we apply Theorem 1.2 with  $m$  replaced by  $k$ ,  $y_k = x^{1-k/d} y^{k/d}$  and  $h_i(s_1, \dots, s_k) = a_i^{(n,k)}(h_k(s_1, \dots, s_k))$ . We derive that

$$\begin{aligned} \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^m} T_i(h(\xi_i)) \right\|_{\mathbb{H}} > x \right) &\leq \sum_{k=d}^m A_k \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{d}} \right) \\ &+ \sum_{k=d}^m B_k \int_1^\infty u (1 + \log u)^{k(k-1)/2-1} \mathbb{P} \left( \sqrt{\sum_{i \in \text{Inc}_k^m} \|a_i^{(n,k)}(h_k(\xi_i^{\text{dec}}))\|_{\mathbb{H}}^2} > \frac{C_k}{m} x^{\frac{d-k}{d}} y^{\frac{k}{d}} u \right) du. \end{aligned}$$

For a convex increasing function  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ , one has

$$\begin{aligned} \mathbb{E} \left[ \varphi \left( \sum_{i \in \text{Inc}_n^m} \|a_i^{(n,k)}(h_k(\xi_i^{\text{dec}}))\|_{\mathbb{H}}^2 \right) \right] &\leq \mathbb{E} \left[ \varphi \left( A_{n,k}^2 \sum_{i \in \text{Inc}_k^m} \frac{\|a_i^{(n,k)}\|_{\mathcal{B}(\mathbb{H})}^2}{A_{n,k}^2} \|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2 \right) \right] \\ &\leq \mathbb{E} \left[ \sum_{i \in \text{Inc}_k^m} \frac{\|a_i^{(n,k)}\|_{\mathcal{B}(\mathbb{H})}^2}{A_{n,k}^2} \varphi \left( A_{n,k}^2 \|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2 \right) \right] \\ &= \mathbb{E} \left[ \varphi \left( A_{n,k}^2 \|h_k(\xi_{[1,k]})\|_{\mathbb{H}}^2 \right) \right] \\ &\leq \mathbb{E} \left[ \varphi \left( A_{n,k}^2 \kappa_m H^2 \right) \right] \\ &\leq \mathbb{E} \left[ \varphi \left( n^{m-k} \sum_{i \in \text{Inc}_n^m} \|T_i\|_{\mathcal{B}(\mathbb{H})}^2 \kappa_m H^2 \right) \right] \\ &\leq \mathbb{E} \left[ \varphi \left( n^{m-d} \sum_{i \in \text{Inc}_n^m} \|T_i\|_{\mathcal{B}(\mathbb{H})}^2 \kappa_m H^2 \right) \right] \end{aligned}$$

hence a use of (A.8) gives

$$\begin{aligned} \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^m} T_i(h(\xi_i)) \right\|_{\mathbb{H}} > x \right) &\leq \sum_{k=d}^m A_k \exp \left( - \left( \frac{x}{y} \right)^{2/d} \right) \\ &+ \sum_{k=d}^m B_k \int_1^\infty \int_1^\infty u (1 + \log u)^{k(k-1)/2-1} \mathbb{P} \left( A_{n,k} \|h_k(\xi_{[1,k]})\|_{\mathbb{H}} > x^{\frac{d-k}{d}} y^{\frac{k}{d}} u \sqrt{v} / (4m) \right) dudv. \end{aligned}$$

We conclude by a use of (3.15).



### 3.4 Proof of the results of Subsection 2.3

Before going into the proofs of each corollary, let us mention the common elements of proof.

Define for  $k \in \llbracket d, m \rrbracket$  and  $\mathbf{i} \in \text{Inc}_n^k$  the random variable

$$Z_{\mathbf{i}}^{(k)} := \sum_{j \in I_k(\mathbf{i})} Z_{n,j},$$

where  $I_k(\mathbf{i})$  denotes the set of indices  $\mathbf{j} = (j_\ell)_{\ell=1}^m \in \text{Inc}_n^m$  such that on a set  $K = \{\ell_1, \dots, \ell_k\}$  of cardinal  $k$ , one has  $j_{\ell_1} = i_1, \dots, j_{\ell_k} = i_k$ . By symmetry of  $h$ , we have

$$\sum_{\mathbf{i} \in \text{Inc}_n^m} Z_{n,\mathbf{i}} h(\xi_{\mathbf{i}}) = \sum_{k=d}^m \sum_{\mathbf{i} \in \text{Inc}_n^k} Z_{n,\mathbf{i}}^{(k)} h_k(\xi_{\mathbf{i}}). \quad (3.16)$$

This decomposition reduces us to show the existence of constants  $a_m, b_m$  and  $c_m$  depending only on  $m$  such that for each  $k \in \llbracket d, m \rrbracket$  and  $x, y$  such that  $x/y > 1$ ,

$$\begin{aligned} \mathbb{P} \left( \left\| \sum_{\mathbf{i} \in \text{Inc}_n^k} Z_{n,\mathbf{i}}^{(k)} h_k(\xi_{\mathbf{i}}) \right\|_{\mathbb{H}} > x \sqrt{N} \sqrt{\min\{N, n^{m-d}\}} \right) &\leq a_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m}} \right) \\ &\quad + b_m \int_1^\infty u (1 + \log u)^{m(m+1)/2} \mathbb{P}(H > c_m y u) du \text{ and} \end{aligned} \quad (3.17)$$

$$\begin{aligned} \mathbb{P} \left( \left\| \sum_{\mathbf{i} \in \text{Inc}_n^k} Z_{n,\mathbf{i}}^{(k)} h_k(\xi_{\mathbf{i}}) \right\|_{\mathbb{H}} > x n^m \sqrt{p_n} \sqrt{\min\{p_n, n^{-d}\}} \right) \\ \leq a_m \exp \left( - \frac{n^m p_n^2}{2} \right) + a_m \exp \left( - \left( \frac{x}{y} \right)^{\frac{2}{m}} \right) + b_m \int_1^\infty u (1 + \log u)^{m(m+1)/2} \mathbb{P}(H > c_m y u) du. \end{aligned} \quad (3.18)$$

*Proof of Corollary 2.6.* By independence between  $(Z_{n,\mathbf{i}})_{\mathbf{i} \in \text{Inc}_n^m}$  and  $(\xi_{\mathbf{i}})_{\mathbf{i} \geq 1}$ , one has

$$\begin{aligned} \mathbb{P} \left( \left\| \sum_{\mathbf{i} \in \text{Inc}_n^k} Z_{n,\mathbf{i}}^{(k)} h_k(\xi_{\mathbf{i}}) \right\|_{\mathbb{H}} > x \right) \\ = \sum_{(z_{\mathbf{i}})_{\mathbf{i} \in \text{Inc}_n^k} \in \{0,1\}^{C_n^k}} \mathbb{P} \left( \left\| \sum_{\mathbf{i} \in \text{Inc}_n^k} \sum_{j \in I_k(\mathbf{i})} z_j h_k(\xi_{\mathbf{i}}) \right\|_{\mathbb{H}} > x \right) \mathbb{P} \left( (Z_{n,\mathbf{i}})_{\mathbf{i} \in \text{Inc}_n^k} = (z_{\mathbf{i}})_{\mathbf{i} \in \text{Inc}_n^k} \right). \end{aligned}$$

Using Theorem 1.2 for fixed  $(z_i)_{i \in \text{Inc}_n^k}$ , we find that for  $y_k$  that will be specified later,

$$\begin{aligned} \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^k} Z_{n,i}^{(k)} h_k(\xi_i) \right\|_{\mathbb{H}} > x \right) &\leq a_k \exp \left( - \left( \frac{x}{y_k} \right)^{\frac{2}{k}} \right) \\ &\quad + b_k \sum_{(z_i)_{i \in \text{Inc}_n^k} \in \{0,1\}^{C_n^k}} \mathbb{P} \left( (Z_{n,i})_{i \in \text{Inc}_n^k} = (z_i)_{i \in \text{Inc}_n^k} \right) \\ &\quad \int_1^\infty u (1 + \log u)^{k(k-1)/2} \mathbb{P} \left( \sum_{i \in \text{Inc}_n^k} \left\| \sum_{j \in I_k(i)} z_j h_k(\xi_i^{\text{dec}}) \right\|_{\mathbb{H}}^2 > u^2 y_k^2 \right) du. \end{aligned}$$

Let  $A := \sum_{i \in \text{Inc}_n^k} \left( \sum_{j \in I_k(i)} z_j \right)^2$ . For a convex increasing function  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ , one has

$$\begin{aligned} \mathbb{E} \left[ \varphi \left( \sum_{i \in \text{Inc}_n^k} \left\| \sum_{j \in I_k(i)} z_j h_k(\xi_i^{\text{dec}}) \right\|_{\mathbb{H}}^2 \right) \right] &= \mathbb{E} \left[ \varphi \left( \frac{1}{A} \sum_{i \in \text{Inc}_n^k} \left( \sum_{j \in I_k(i)} z_j \right)^2 A \|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2 \right) \right] \\ &\leq \mathbb{E} \left[ \frac{1}{A} \sum_{i \in \text{Inc}_n^k} \left( \sum_{j \in I_k(i)} z_j \right)^2 \varphi \left( A \|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2 \right) \right] \\ &\leq \mathbb{E} \left[ \frac{1}{A} \sum_{i \in \text{Inc}_n^k} \left( \sum_{j \in I_k(i)} z_j \right)^2 \varphi \left( A \kappa_m H^2 \right) \right] \end{aligned}$$

hence  $\sum_{i \in \text{Inc}_n^k} \left\| \sum_{j \in I_k(i)} z_j h_k(\xi_i) \right\|_{\mathbb{H}}^2 \stackrel{\text{conv}}{\leq} A \kappa_m H^2$ . Since  $\sqrt{\sum c_i} \leq \sum \sqrt{c_i}$ , one has  $A \leq N^2$ . Moreover, by Cauchy-Schwarz,  $A \leq n^{m-k} N \leq n^{m-d} N$ . Overall,

$$\sum_{i \in \text{Inc}_n^k} \left\| \sum_{j \in I_k(i)} z_j h_k(\xi_i^{\text{dec}}) \right\|_{\mathbb{H}}^2 \stackrel{\text{conv}}{\leq} N \min \{N, n^{m-d}\} \kappa_m H^2$$

hence by (A.8) and (3.15), we get

$$\begin{aligned} \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^k} Z_{n,i}^{(k)} h_k(\xi_i) \right\|_{\mathbb{H}} > x \right) &\leq a_k \exp \left( - \left( \frac{x}{y_k} \right)^{2/k} \right) \\ &\quad + b_k \int_1^\infty u (1 + \log u)^{k(k+1)/2} \mathbb{P} \left( \sqrt{N \min \{N, n^{m-d}\}} \kappa'_m H > y_k u \right) du. \end{aligned}$$

Bounding the exponent  $k(k+1)/2$  by  $m(m+1)/2$ , choosing  $y_k$  such that  $\left(\frac{x}{y_k}\right)^{2/k} = \left(\frac{x}{y}\right)^{2/m}$  and using the fact that  $x/y > 1$ , one gets

$$\begin{aligned} \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^k} Z_{n,i}^{(k)} h_k(\xi_i) \right\|_{\mathbb{H}} > x \right) &\leq a_k \exp \left( - \left( \frac{x}{y} \right)^{2/m} \right) \\ &\quad + b_k \int_1^\infty u (1 + \log u)^{m(m+1)/2} \mathbb{P} \left( \sqrt{N \min \{N, n^{m-d}\}} \kappa'_m H > y u \right) du \end{aligned}$$

from which we derive (3.17) by applying the previous inequality with  $x$  and  $y$  replaced respectively by  $x (N \min \{N, n^{m-d}\})^{-1/2}$  and  $y (N \min \{N, n^{m-d}\})^{-1/2}$ .  $\square$

*Proof of Corollary 2.7.* We start from

$$\mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^k} Z_{n,i}^{(k)} h_k(\xi_i) \right\|_{\mathbb{H}} > xn^m \sqrt{p_n} \sqrt{\min \{p_n, n^{-d}\}} \right) \leq P_1 + P_2,$$

where

$$P_1 := \mathbb{P} \left( \sum_{i \in \text{Inc}_n^m} Z_{n,i} > 2n^m p_n \right) \text{ and}$$

$$P_2 := \mathbb{P} \left( \left\{ \left\| \sum_{i \in \text{Inc}_n^k} Z_{n,i}^{(k)} h_k(\xi_i) \right\|_{\mathbb{H}} > xn^m \sqrt{p_n} \sqrt{\min \{p_n, n^{-d}\}} \right\} \cap \left\{ \sum_{i \in \text{Inc}_n^m} Z_{n,i} \leq 2n^m p_n \right\} \right).$$

The term  $P_1$  is bounded, via Azuma-Hoeffding's inequality, by  $2 \exp(-n^m p_n^2/2)$ . Using independence between  $(Z_{n,i})_{i \in \text{Inc}_n^m}$  and  $(\xi_i)_{i \geq 1}$ , we get

$$P_2 = \sum_{(z_i)_{i \in \text{Inc}_n^k} \in \{0,1\}^{C_n^k}} \mathbb{P} \left( \left\| \sum_{i \in \text{Inc}_n^k} \sum_{j \in I_k(i)} z_j h_k(\xi_i) \right\|_{\mathbb{H}} > xn^m \sqrt{p_n} \sqrt{\min \{p_n, n^{-d}\}} \right) \\ \mathbf{1}_{\sum_{i \in \text{Inc}_n^m} z_i \leq 2n^m p_n} \mathbb{P} \left( (Z_{n,i})_{i \in \text{Inc}_n^k} = (z_i)_{i \in \text{Inc}_n^k} \right).$$

We apply Theorem 1.2 for fixed  $(z_i)_{i \in \text{Inc}_n^k}$  and get

$$P_2 \leq a_k \exp \left( - \left( \frac{x}{y_k} \right)^{2/k} \right) + b_k \int_1^\infty u (1 + \log u)^{\frac{k(k-1)}{2}} \sum_{(z_i)_{i \in \text{Inc}_n^k} \in \{0,1\}^{C_n^k}} \mathbb{P} \left( (Z_{n,i})_{i \in \text{Inc}_n^k} = (z_i)_{i \in \text{Inc}_n^k} \right) \\ \mathbb{P} \left( \sum_{i \in \text{Inc}_n^k} \left\| \sum_{j \in I_k(i)} z_j h_k(\xi_i^{\text{dec}}) \right\|_{\mathbb{H}}^2 > u^2 (xn^m)^2 p_n \min \{p_n, n^{-d}\} \right) \mathbf{1}_{\sum_{i \in \text{Inc}_n^m} z_i \leq 2n^m p_n} du.$$

For  $(z_i)_{i \in \text{Inc}_n^m}$  such that  $\sum_{i \in \text{Inc}_n^m} z_i \leq 2n^m p_n$  and a convex non-decreasing function  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ , letting  $A := \sum_{i \in \text{Inc}_n^k} \left( \sum_{j \in I_k(i)} z_j \right)^2$ , one has

$$\mathbb{E} \left[ \varphi \left( \sum_{i \in \text{Inc}_n^k} \left\| \sum_{j \in I_k(i)} z_j h_k(\xi_i^{\text{dec}}) \right\|_{\mathbb{H}}^2 \right) \right] = \mathbb{E} \left[ \varphi \left( \frac{1}{A} \sum_{i \in \text{Inc}_n^k} \left( \sum_{j \in I_k(i)} z_j \right)^2 A \|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2 \right) \right] \\ \leq \mathbb{E} \left[ \frac{1}{A} \sum_{i \in \text{Inc}_n^k} \left( \sum_{j \in I_k(i)} z_j \right)^2 \varphi \left( A \|h_k(\xi_i^{\text{dec}})\|_{\mathbb{H}}^2 \right) \right] \\ \leq \mathbb{E} [\varphi(A \kappa_m H^2)].$$

Since  $\sum_{i \in \text{Inc}_n^m} z_i \leq 2np_n$ , an application of  $\sqrt{\sum c_i} \leq \sum \sqrt{c_i}$  gives that  $A^2 \leq 2n^m p_n$ . Moreover, convexity of the square gives  $A \leq n^{m-k} 2n^m p_n \leq 2n^{2m-d} p_n$ . The combination of the obtained estimates gives  $\sum_{i \in \text{Inc}_n^k} \left\| \sum_{j \in I_k(i)} z_j h_k(\xi_i^{\text{dec}}) \right\|_{\mathbb{H}}^2 \leq \kappa_m \min \{ \sqrt{2n^m p_n}, 2n^{2m-d} p_n \} H^2$ . We then conclude similarly as in the proof of Corollary 2.6: we use (A.8), then (3.15) and take  $y_k$  making the terms  $\exp \left( - \left( \frac{x}{y_k} \right)^{2/k} \right)$ ,  $k \in \llbracket d, m \rrbracket$ , equal and finally, the fact that  $x/y$  is bigger than one allows to find a bound independent of  $k$ . This ends the proof of Corollary 2.7.  $\square$

## A A deviation inequality for Hilbert-valued martingale differences

### A.1 Real-valued case

In this section, we collect existing results in the literature in order to derive an exponential inequality for real-valued martingale difference sequence. The tail of the maxima of absolute values of a martingale are controlled via an exponential term and the tail of the sum of squares. This will be the starting point

Fan et al. (2015) showed the following.

**Theorem A.1** (Theorem 2.1 in Fan et al. (2015)). *Let  $(D_j)_{j \geq 1}$  be a real-valued martingale difference sequence with respect to the filtration  $(\mathcal{F}_j)_{j \geq 0}$ . Suppose that there exists random variables  $V_{j-1}$ ,  $j \in \llbracket 1, n \rrbracket$ , which are non-negative and  $\mathcal{F}_{j-1}$ -measurable, non-negative functions  $f$  and  $g$  such that for some positive  $\lambda$  and for all  $j \in \llbracket 1, n \rrbracket$ ,*

$$\mathbb{E} \left[ \exp \left( \lambda D_j - g(\lambda) D_j^2 \right) \mid \mathcal{F}_{j-1} \right] \leq 1 + f(\lambda) V_{j-1}. \quad (\text{A.1})$$

Then for all  $x, v, w > 0$ ,

$$\mathbb{P} \left( \bigcup_{k=1}^n \left\{ \sum_{j=1}^k D_j \geq x \right\} \cap \left\{ \sum_{j=1}^k D_j^2 \leq v^2 \right\} \cap \left\{ \sum_{j=1}^k V_{j-1} \leq w \right\} \right) \leq \exp \left( -\lambda x + g(\lambda) v^2 + f(\lambda) w \right). \quad (\text{A.2})$$

Notice that the condition (A.1) is satisfied for all  $\lambda > 0$  when  $f(\lambda) = g(\lambda) = \lambda^2/2$  and  $V_{j-1} = \mathbb{E} [D_j^2 \mid \mathcal{F}_{j-1}]$  provided that  $\mathbb{E} [D_j^2]$  is finite for each  $j$ . After having chosen  $w = v^2$ , optimized over  $\lambda$  the right hand side of (A.2) and applying Theorem A.1 to  $D_j$  and then to  $-D_j$ , we derive that for all  $x$  and  $y > 0$

$$\mathbb{P} \left( \max_{1 \leq k \leq n} \left| \sum_{j=1}^k D_j \right| > x \right) \leq 2 \exp \left( -\frac{x^2}{y^2} \right) + \mathbb{P} \left( \sum_{j=1}^n (D_j^2 + \mathbb{E} [D_j^2 \mid \mathcal{F}_{j-1}]) > \frac{y^2}{2} \right). \quad (\text{A.3})$$

An inequality in this spirit has been established by Bercu and Touati (2008), Theorem 2.1, without the max but the event involving the sum of squares and conditional variances is in the probability in the left hand side.

## A.2 Dimension reduction in Hilbert spaces

An extension of (A.2) to the Hilbert-valued case requires the following dimension reduction result. This was done by [Kallenberg and Sztencel \(1991\)](#) in the continuous time case, and in Proposition 5.8.3 in [Kwapień and Woyczyński \(1992\)](#) in the discrete case.

**Lemma A.2.** *Let  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  be a separable Hilbert space, let  $\|x\|_{\mathbb{H}} := \sqrt{\langle x, x \rangle}$  and let  $(D_j)_{j \geq 1}$  be a  $\mathbb{H}$ -valued martingale difference sequence with respect to the filtration  $(\mathcal{F}_j)_{j \geq 0}$ . Enlarging the probability space if needed, there exists an  $\mathbb{R}^2$ -valued martingale difference sequence  $((d_{1,j}, d_{2,j}))_{j \geq 1}$  with respect to the filtration  $(\mathcal{G}_j)_{j \geq 0}$  such that for each  $j \geq 1$ ,  $\|D_j\|_{\mathbb{H}}^2 = d_{1,j}^2 + d_{2,j}^2$  and for each  $n \geq 1$ ,*

$$\sum_{j=1}^n \mathbb{E} [\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_{j-1}] = \sum_{j=1}^n \mathbb{E} [d_{j,1}^2 \mid \mathcal{G}_{j-1}] + \sum_{j=1}^n \mathbb{E} [d_{j,2}^2 \mid \mathcal{G}_{j-1}] \quad \text{and} \quad (\text{A.4})$$

$$\left\| \sum_{j=1}^n D_j \right\|_{\mathbb{H}}^2 = \left( \sum_{j=1}^n d_{1,j} \right)^2 + \left( \sum_{j=1}^n d_{2,j} \right)^2. \quad (\text{A.5})$$

A consequence of such a dimension reduction is that a deviation inequality for real valued martingale difference sequences in terms of tails of norms of increments can be converted into a similar inequality for a Hilbert-valued martingale difference sequences. An inequality in the spirit of (A.3) reads as follows.

**Proposition A.3.** *For each separable Hilbert space  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  and each  $\mathbb{H}$ -valued martingale difference sequence  $(D_j)_{j \geq 1}$  with respect to a filtration  $(\mathcal{F}_j)_{j \geq 0}$ ,*

$$\mathbb{P} \left( \max_{1 \leq k \leq n} \left\| \sum_{j=1}^k D_j \right\|_{\mathbb{H}} > x \right) \leq 4 \exp \left( -\frac{x^2}{y^2} \right) + 2 \mathbb{P} \left( \sum_{j=1}^n (\|D_j\|_{\mathbb{H}}^2 + \mathbb{E} [\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_{j-1}]) > \frac{y^2}{8} \right), \quad (\text{A.6})$$

where  $\|x\|_{\mathbb{H}} = \sqrt{\langle x, x \rangle}$ .

*Proof.* Let  $\mathbb{H}$  be a separable Hilbert space and let  $(D_j)_{j \geq 1}$  be a  $\mathbb{H}$ -valued martingale difference sequence with respect to a filtration  $(\mathcal{F}_j)_{j \geq 0}$ . By Lemma A.2, there exist an  $\mathbb{R}^2$ -valued martingale difference sequence  $((d_{1,j}, d_{2,j}))_{j \geq 1}$  such that for each  $j \geq 1$ ,  $\|D_j\|_{\mathbb{H}}^2 = d_{1,j}^2 + d_{2,j}^2$  and (A.4), (A.5) hold for each  $n$ . In particular,

$$\left\| \sum_{j=1}^n D_j \right\|_{\mathbb{H}} \leq \left| \sum_{j=1}^n d_{1,j} \right| + \left| \sum_{j=1}^n d_{2,j} \right| \quad (\text{A.7})$$

from which the following inequality follows:

$$\mathbb{P} \left( \max_{1 \leq k \leq n} \left\| \sum_{j=1}^k D_j \right\|_{\mathbb{H}} > x \right) \leq \mathbb{P} \left( \max_{1 \leq k \leq n} \left| \sum_{j=1}^k d_{1,j} \right| > \frac{x}{2} \right) + \mathbb{P} \left( \max_{1 \leq k \leq n} \left| \sum_{j=1}^k d_{2,j} \right| > \frac{x}{2} \right).$$

An application of (A.3) with  $y^2$  replaced by  $y_1^2 = y^2/4$  gives

$$\begin{aligned} \mathbb{P} \left( \max_{1 \leq k \leq n} \left\| \sum_{j=1}^k D_j \right\|_{\mathbb{H}} > x \right) &\leq 4 \exp \left( -\frac{x^2}{y^2} \right) \\ &+ \mathbb{P} \left( \sum_{j=1}^n d_{1,j}^2 + \mathbb{E} [d_{1,j}^2 \mid \mathcal{G}_{j-1}] > \frac{y^2}{8} \right) + \mathbb{P} \left( \sum_{j=1}^n d_{2,j}^2 + \mathbb{E} [d_{2,j}^2 \mid \mathcal{G}_{j-1}] > \frac{y^2}{8} \right) \end{aligned}$$

and (A.6) follows from the combination of (A.4) and (A.5).  $\square$

The following ordering was studied in Rüschendorf (1981).

**Definition A.4.** Let  $X$  and  $Y$  be two real-valued random variables. We say that  $X \leq_{\text{conv}} Y$  if for each nondecreasing convex  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  such that the expectations  $\mathbb{E}[\varphi(X)]$  and  $\mathbb{E}[\varphi(Y)]$  exist,  $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(Y)]$ .

The point of this ordering is that if  $X \leq_{\text{conv}} Y$ , one can formulate a tail inequality for  $X$  in terms of tails of  $Y$ . More precisely, Lemma 2.1 in Giraud (2021) gives that if  $X$  and  $Y$  are nonnegative random variables such that  $X \leq_{\text{conv}} Y$ , then for each positive  $t$ ,

$$\mathbb{P}(X > t) \leq \int_1^\infty \mathbb{P}\left(Y > t \frac{v}{4}\right) dv. \quad (\text{A.8})$$

When the conditional variances admit a particular form, it is possible to express the tails of maxima of a martingale by the sum of squares of the norm of the increments, that is, without conditional moment.

**Proposition A.5.** Let  $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}})$  be a separable Hilbert space. Let  $(D_j)_{j \geq 1}$  be a square integrable martingale difference sequence with respect to a filtration  $(\mathcal{F}_j)_{j \geq 0}$ . Suppose that

$$\mathbb{E} [\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_{j-1}] = \mathbb{E} [\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_0] \quad a.s..$$

Then for each positive  $x$  and  $y$ , the following inequality holds:

$$\mathbb{P} \left( \max_{1 \leq k \leq n} \left\| \sum_{j=1}^k D_j \right\|_{\mathbb{H}} > x \right) \leq 4 \exp \left( -\frac{x^2}{y^2} \right) + 4 \int_1^\infty u \mathbb{P} \left( \sqrt{\sum_{j=1}^n \|D_j\|_{\mathbb{H}}^2} > y \frac{u}{8} \right) du. \quad (\text{A.9})$$

*Proof.* We combine (A.6) with (A.8), applied with  $X = \sum_{j=1}^n (\|D_j\|_{\mathbb{H}}^2 + \mathbb{E} [\|D_j\|_{\mathbb{H}}^2 \mid \mathcal{F}_{j-1}])$  and  $Y = 2 \sum_{j=1}^n \|D_j\|_{\mathbb{H}}^2$ .  $\square$

## B Decoupling inequalities for $U$ -statistics

Let  $U_n$  be a  $U$ -statistic defined as in (1.1). Suppose for simplicity that  $h_i = h$ . In general, its treatment is a difficult problem because the contribution of each random variable  $\xi_i$  appears

in several coordinates of the function  $h$  and the random field defined by  $X_{\mathbf{i}} = h(\xi_{\mathbf{i}})$  is not stationary. Moreover, the involved filtrations are not easy to deal with. A much easier object is the decoupled  $U$ -statistic  $U_n^{\text{dec}}$  for which the entries are replaced by independent copies of  $(\xi_i)_{i \geq 1}$ . More precisely, we define

$$U_n^{\text{dec}}((h_{\mathbf{i}})) := \sum_{\mathbf{i} \in \text{Inc}_n^m} h_{\mathbf{i}}(\xi_{\mathbf{i}}^{\text{dec}}), \quad (\text{B.1})$$

$\xi_{\mathbf{i}}^{\text{dec}} = (\xi_{i_1}^{(1)}, \dots, \xi_{i_m}^{(m)})$ , and the sequences  $(\xi_i^{(1)})_{i \geq 1}, \dots, (\xi_i^{(m)})_{i \geq 1}$ , are mutually independent and have the same distribution as the sequence  $(\xi_i)_{i \geq 1}$ .

An important feature of decoupled  $U$ -statistics is that the tail of the  $U$ -statistic can be controlled in terms of that of its decoupled version.

**Proposition B.1** (de la Peña and Montgomery-Smith (1995)). *For each  $m \geq 1$ , there exists a constant  $K_m$  depending only on  $m$  such that if  $(\xi_i)_{i \geq 1}$  is an i.i.d. sequence taking values in a measurable space  $(S, \mathcal{S})$ ,  $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$  a separable Banach space,  $h_{\mathbf{i}}: S^m \rightarrow \mathbb{B}$  and let  $U_n((h_{\mathbf{i}}))$  and  $U_n^{\text{dec}}((h_{\mathbf{i}}))$  are defined respectively as in (1.1) and (B.1), then the following inequality holds for each positive  $x$ :*

$$\mathbb{P}(\|U_n((h_{\mathbf{i}}))\|_{\mathbb{B}} > x) \leq K_m \mathbb{P}(K_m \|U_n^{\text{dec}}((h_{\mathbf{i}}))\|_{\mathbb{B}} > x).$$

Since we are interested in maxima of  $U$ -statistics, we will need the following consequence of Proposition B.1.

**Corollary B.2.** *For each  $m \geq 1$ , there exists a constant  $K_m$  depending only on  $m$  such that if  $(\xi_i)_{i \geq 1}$  is an i.i.d. sequence taking values in a measurable space  $(S, \mathcal{S})$ ,  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$  a separable Hilbert space,  $h_{\mathbf{i}}: S^m \rightarrow \mathbb{H}$  and let  $U_n((h_{\mathbf{i}}))$  and  $U_n^{\text{dec}}((h_{\mathbf{i}}))$  are defined respectively as in (1.1) and (B.1), then the following inequality holds for each positive  $x$  and  $N \geq m$ :*

$$\mathbb{P}\left(\max_{m \leq n \leq N} \|U_n((h_{\mathbf{i}}))\|_{\mathbb{H}} > x\right) \leq K_m \mathbb{P}\left(K_m \max_{m \leq n \leq N} \|U_n^{\text{dec}}((h_{\mathbf{i}}))\|_{\mathbb{H}} > x\right).$$

*Proof.* We apply Proposition B.1 to the following setting: we take  $\mathbb{B} = \mathbb{H}^N$  endowed with the norm  $\|(y_n)_{n=1}^N\|_{\mathbb{B}} := \max_{1 \leq n \leq N} \|y_n\|_{\mathbb{H}}$  and for  $\mathbf{i} \in \text{Inc}^m$  consider  $\tilde{h}_{\mathbf{i}}: S^m \rightarrow \mathbb{H}^N$  defined as

$$\tilde{h}_{\mathbf{i}}(s_1, \dots, s_m) = \left(H_{\mathbf{i}}^{(n)}(s_1, \dots, s_m)\right)_{n=1}^N,$$

where

$$H_{\mathbf{i}}^{(n)}(s_1, \dots, s_m) = h_{\mathbf{i}\mathbf{i}}(s_1, \dots, s_m) \mathbf{1}_{i_m \leq n}.$$

In this way,

$$\max_{m \leq n \leq N} \|U_n((h_{\mathbf{i}}))\|_{\mathbb{H}} = \left\| U_N\left(\left(\tilde{h}_{\mathbf{i}}\right)\right) \right\|_{\mathbb{B}}$$

and a similar equality holds for the decoupled versions of the  $U$ -statistics.  $\square$

## References

- Radosław Adamczak. Moment inequalities for  $U$ -statistics. *Ann. Probab.*, 34(6): 2288–2314, 2006. ISSN 0091-1798. doi: 10.1214/009117906000000476. URL <https://doi.org/10.1214/009117906000000476>.
- Miguel A. Arcones. A Bernstein-type inequality for  $U$ -statistics and  $U$ -processes. *Statist. Probab. Lett.*, 22(3):239–247, 1995. ISSN 0167-7152,1879-2103. doi: 10.1016/0167-7152(94)00072-G. URL [https://doi.org/10.1016/0167-7152\(94\)00072-G](https://doi.org/10.1016/0167-7152(94)00072-G).
- Bernard Bercu and Abderrahmen Touati. Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.*, 18(5):1848–1869, 2008. ISSN 1050-5164. doi: 10.1214/07-AAP506. URL <https://doi.org/10.1214/07-AAP506>.
- Peter J. Bickel and Ya’acov Ritov. An exponential inequality for  $U$ -statistics with applications to testing. *Probab. Engrg. Inform. Sci.*, 9(1):39–52, 1995. ISSN 0269-9648,1469-8951. doi: 10.1017/S0269964800003673. URL <https://doi.org/10.1017/S0269964800003673>.
- Gunnar Blom. Some properties of incomplete  $U$ -statistics. *Biometrika*, 63(3):573–580, 1976. ISSN 0006-3444,1464-3510. doi: 10.1093/biomet/63.3.573. URL <https://doi.org/10.1093/biomet/63.3.573>.
- I. S. Borisov. Exponential inequalities for the distributions of von Mises and  $U$ -statistics. In *Probability theory and mathematical statistics, Vol. I (Vilnius, 1989)*, pages 166–178. “Mokslas”, Vilnius, 1990. ISBN 90-6764-128-6.
- I. S. Borisov and A. A. Bystrov. Exponential inequalities for the distributions of canonical multiple partial sum processes. *Teor. Veroyatn. Primen.*, 64(2):209–227, 2019. ISSN 0040-361X,2305-3151. doi: 10.1137/S0040585X97T98943X. URL <https://doi.org/10.1137/S0040585X97T98943X>.
- I. S. Borisov and N. V. Volodko. Exponential inequalities for the distributions of canonical  $U$ - and  $V$ -statistics of dependent observations [translation of mr2500124]. *Siberian Adv. Math.*, 19(1):1–12, 2009. ISSN 1055-1344,1934-8126. doi: 10.3103/S1055134409010015. URL <https://doi.org/10.3103/S1055134409010015>.
- Anirvan Chakraborty and Probal Chaudhuri. A Wilcoxon-Mann-Whitney-type test for infinite-dimensional data. *Biometrika*, 102(1):239–246, 2015. ISSN 0006-3444,1464-3510. doi: 10.1093/biomet/asu072. URL <https://doi.org/10.1093/biomet/asu072>.
- Anirvan Chakraborty and Probal Chaudhuri. Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann. Statist.*, 45(2):771–799, 2017. ISSN 0090-5364,2168-8966. doi: 10.1214/16-AOS1467. URL <https://doi.org/10.1214/16-AOS1467>.



- Xiaohui Chen and Kengo Kato. Randomized incomplete  $U$ -statistics in high dimensions. *Ann. Statist.*, 47(6):3127–3156, 2019. ISSN 0090-5364,2168-8966. doi: 10.1214/18-AOS1773. URL <https://doi.org/10.1214/18-AOS1773>.
- Tasos C. Christofides. Probability inequalities with exponential bounds for  $U$ -statistics. *Statist. Probab. Lett.*, 12(3):257–261, 1991. ISSN 0167-7152,1879-2103. doi: 10.1016/0167-7152(91)90087-8. URL [https://doi.org/10.1016/0167-7152\(91\)90087-8](https://doi.org/10.1016/0167-7152(91)90087-8).
- Stéphan Cléménçon, Gábor Lugosi, and Nicolas Vayatis. Ranking and empirical minimization of  $U$ -statistics. *Ann. Statist.*, 36(2):844–874, 2008. ISSN 0090-5364,2168-8966. doi: 10.1214/009052607000000910. URL <https://doi.org/10.1214/009052607000000910>.
- Stephan Cléménçon, Igor Colin, and Aurélien Bellet. Scaling-up empirical risk minimization: optimization of incomplete  $U$ -statistics. *J. Mach. Learn. Res.*, 17:Paper No. 76, 36, 2016. ISSN 1532-4435,1533-7928.
- Victor H. de la Peña and S. J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate  $U$ -statistics. *Ann. Probab.*, 23(2):806–816, 1995. ISSN 0091-1798. URL [http://links.jstor.org/sici?sici=0091-1798\(199504\)23:2<806:DIFTTP>2.0.CO;2-H&origin=MSN](http://links.jstor.org/sici?sici=0091-1798(199504)23:2<806:DIFTTP>2.0.CO;2-H&origin=MSN).
- Herold Dehling, Kata Vuk, and Martin Wendler. Change-point detection based on weighted two-sample  $U$ -statistics. *Electron. J. Stat.*, 16(1):862–891, 2022. ISSN 1935-7524. doi: 10.1214/21-ejs1964. URL <https://doi.org/10.1214/21-ejs1964>.
- Jian Ding, James R. Lee, and Yuval Peres. Markov type and threshold embeddings. *Geom. Funct. Anal.*, 23(4):1207–1229, 2013. ISSN 1016-443X. doi: 10.1007/s00039-013-0234-7. URL <https://doi.org/10.1007/s00039-013-0234-7>.
- Quentin Duchemin, Yohann De Castro, and Claire Lacour. Concentration inequality for  $U$ -statistics of order two for uniformly ergodic Markov chains. *Bernoulli*, 29(2):929–956, 2023. ISSN 1350-7265,1573-9759. doi: 10.3150/22-bej1485. URL <https://doi.org/10.3150/22-bej1485>.
- Alexander Dürre and Davy Paindaveine. On the consistency of incomplete  $U$ -statistics under infinite second-order moments. *Statist. Probab. Lett.*, 193:Paper No. 109714, 8, 2023. ISSN 0167-7152,1879-2103. doi: 10.1016/j.spl.2022.109714. URL <https://doi.org/10.1016/j.spl.2022.109714>.
- Xiequan Fan, Ion Grama, and Quansheng Liu. Exponential inequalities for martingales with applications. *Electron. J. Probab.*, 20:no. 1, 22, 2015. doi: 10.1214/EJP.v20-3496. URL <https://doi.org/10.1214/EJP.v20-3496>.

- Evarist Giné, Rafał Łatała, and Joel Zinn. Exponential and moment inequalities for  $U$ -statistics. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 13–38. Birkhäuser Boston, Boston, MA, 2000. ISBN 0-8176-4160-2.
- Davide Giraud. An exponential inequality for  $U$ -statistics of I.I.D. data. *Teor. Veroyatn. Primen.*, 66(3):508–533, 2021. ISSN 0040-361X. doi: 10.4213/tvp5392. URL <https://doi.org/10.4213/tvp5392>.
- Davide Giraud. Deviation inequality for Banach-valued orthomartingales. *Stochastic Process. Appl.*, 175:Paper No. 104391, 20, 2024a. ISSN 0304-4149,1879-209X. doi: 10.1016/j.spa.2024.104391. URL <https://doi.org/10.1016/j.spa.2024.104391>.
- Davide Giraud. Deviation and moment inequalities for banach-valued  $u$ -statistics. 2024b. doi: 10.48550/arXiv.2405.01902. URL <https://arxiv.org/abs/2405.01902>.
- Fang Han. An exponential inequality for  $U$ -statistics under mixing conditions. *J. Theoret. Probab.*, 31(1):556–578, 2018. ISSN 0894-9840,1572-9230. doi: 10.1007/s10959-016-0722-4. URL <https://doi.org/10.1007/s10959-016-0722-4>.
- Michel Harel, Colm Art O’Cinneide, and Madan L. Puri. Weak convergence of weighted empirical  $U$ -statistics processes for dependent random variables. *J. Multivariate Anal.*, 48(2):297–314, 1994. ISSN 0047-259X,1095-7243. doi: 10.1006/jmva.1994.1007. URL <https://doi.org/10.1006/jmva.1994.1007>.
- Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948. ISSN 0003-4851. doi: 10.1214/aoms/1177730196. URL <https://doi.org/10.1214/aoms/1177730196>.
- Christian Houdré and Patricia Reynaud-Bouret. Exponential inequalities, with constants, for  $U$ -statistics of order two. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 55–69. Birkhäuser, Basel, 2003. ISBN 3-7643-2197-0.
- Junwei Hu and Lihong Wang. A weighted  $U$ -statistic based change point test for multivariate time series. *Statist. Papers*, 64(3):753–778, 2023. ISSN 0932-5026,1613-9798. doi: 10.1007/s00362-022-01341-9. URL <https://doi.org/10.1007/s00362-022-01341-9>.
- Wen Juan Hu, Liang Wang, Bao Xue Zhang, and Guo Chang Wang. A two sample test based on  $U$ -statistic for functional data. *Acta Math. Sin. (Engl. Ser.)*, 39(3):533–552, 2023. ISSN 1439-8516,1439-7617. doi: 10.1007/s10114-023-1608-x. URL <https://doi.org/10.1007/s10114-023-1608-x>.
- Svante Janson. The asymptotic distributions of incomplete  $U$ -statistics. *Z. Wahrsch. Verw. Gebiete*, 66(4):495–505, 1984. ISSN 0044-3719. doi: 10.1007/BF00531887. URL <https://doi.org/10.1007/BF00531887>.

- Feiyu Jiang, Runmin Wang, and Xiaofeng Shao. Robust inference for change points in high dimension. *J. Multivariate Anal.*, 193:Paper No. 105114, 23, 2023. ISSN 0047-259X,1095-7243. doi: 10.1016/j.jmva.2022.105114. URL <https://doi.org/10.1016/j.jmva.2022.105114>.
- Olav Kallenberg and Rafał Sztencel. Some dimension-free features of vector-valued martingales. *Probab. Theory Related Fields*, 88(2):215–247, 1991. ISSN 0178-8051,1432-2064. doi: 10.1007/BF01212560. URL <https://doi.org/10.1007/BF01212560>.
- Stanisław Kwapien and Wojbor A. Woyczyński. *Random series and stochastic integrals: single and multiple*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. ISBN 0-8176-3572-6. doi: 10.1007/978-1-4612-0425-1. URL <https://doi.org/10.1007/978-1-4612-0425-1>.
- Matthias Löwe and Sara Terveer. A central limit theorem for incomplete  $U$ -statistics over triangular arrays. *Braz. J. Probab. Stat.*, 35(3):499–522, 2021. ISSN 0103-0752,2317-6199. doi: 10.1214/20-BJPS492. URL <https://doi.org/10.1214/20-BJPS492>.
- Sijie Luo. On Azuma-type inequalities for Banach space-valued martingales. *J. Theoret. Probab.*, 35(2):772–800, 2022. ISSN 0894-9840,1572-9230. doi: 10.1007/s10959-021-01086-5. URL <https://doi.org/10.1007/s10959-021-01086-5>.
- Andreas Maurer. Exponential finite sample bounds for incomplete  $u$ -statistics. 2022. doi: 10.48550/arXiv.2207.03136.
- Masoud M. Nasari. Strong law of large numbers for weighted  $U$ -statistics: application to incomplete  $U$ -statistics. *Statist. Probab. Lett.*, 82(6):1208–1217, 2012. ISSN 0167-7152,1879-2103. doi: 10.1016/j.spl.2012.03.010. URL <https://doi.org/10.1016/j.spl.2012.03.010>.
- Bouameur Ragbi. Weak convergence of weighted multivariate empirical  $U$ -statistics processes under mixing condition. *J. Statist. Plann. Inference*, 61(1):1–16, 1997. ISSN 0378-3758,1873-1171. doi: 10.1016/S0378-3758(96)00147-4. URL [https://doi.org/10.1016/S0378-3758\(96\)00147-4](https://doi.org/10.1016/S0378-3758(96)00147-4).
- Bouameur Ragbi. Weak convergence for rectangle-indexed weighted multivariate empirical  $U$ -statistic processes under mixing conditions. *J. Math. Anal. Appl.*, 221(2):476–498, 1998. ISSN 0022-247X,1096-0813. doi: 10.1006/jmaa.1997.5894. URL <https://doi.org/10.1006/jmaa.1997.5894>.
- Ludger Rüschendorf. Ordering of distributions and rearrangement of functions. *Ann. Probab.*, 9(2):276–283, 1981. ISSN 0091-1798. URL [http://links.jstor.org/sici?sici=0091-1798\(198104\)9:2<276:OODARO>2.0.CO;2-#&origin=MSN](http://links.jstor.org/sici?sici=0091-1798(198104)9:2<276:OODARO>2.0.CO;2-#&origin=MSN).
- F. H. Ruymgaart and M. C. A. van Zuijlen. Empirical  $U$ -statistics processes. *J. Statist. Plann. Inference*, 32(2):259–269, 1992. ISSN 0378-3758,1873-1171. doi: 10.1016/0378-3758(92)90051-S. URL [https://doi.org/10.1016/0378-3758\(92\)90051-S](https://doi.org/10.1016/0378-3758(92)90051-S).

- Yandi Shen, Fang Han, and Daniela Witten. Exponential inequalities for dependent V-statistics via random Fourier features. *Electron. J. Probab.*, 25:Paper No. 7, 18, 2020. ISSN 1083-6489. doi: 10.1214/20-ejp411. URL <https://doi.org/10.1214/20-ejp411>.
- Qiyang Wang. Bernstein type inequalities for degenerate  $U$ -statistics with applications. *Chinese Ann. Math. Ser. B*, 19(2):157–166, 1998. ISSN 0252-9599,1860-6261. A Chinese summary appears in Chinese Ann. Math. Ser. A **19** (1998), no. 2, 283.
- Lea Wegner and Martin Wendler. Robust change-point detection for functional time series based on  $U$ -statistics and dependent wild bootstrap, 2023. URL <https://arxiv.org/abs/2206.01458>.