Tail Bounds for Canonical U-Statistics and U-Processes with Unbounded Kernels*

Abhishek Chakrabortty† and Arun Kumar Kuchibhotla‡

November 10, 2018

Abstract

In this paper, we prove exponential tail bounds for canonical (or degenerate) U-statistics and U-processes under exponential-type tail assumptions on the kernels. Most of the existing results in the relevant literature often assume bounded kernels or obtain sub-optimal tail behavior under unbounded kernels. We obtain sharp rates and optimal tail behavior under sub-Weibull kernel functions. Some examples from non-parametric regression literature are considered.

Keywords and phrases: Degenerate U-statistics and U-processes, Unbounded kernels, Sub-Weibull tails, Exponential tail bounds, Non-parametric regression.

1 Introduction and Motivation

In this paper, we study the properties of the U-statistic

$$U_n = \sum_{1\le i\neq j\le n} \phi(Y_i)w_{i,j}(X_i, X_j)\psi(Y_j), \tag{1}$$

where $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$ are independent but possibly non-identically distributed random variables defined on a domain $\mathfrak{X}\times\mathfrak{Y}$, for functions $\phi, \psi : \mathfrak{Y} \to \mathbb{R}$ and $\{w_{i,j} : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R} : 1 \le i \neq j \le n\}$ represents a set of bivariate functions. U-statistics of type (1) appears frequently in nonparametric statistics literature although not exlcusively. For example, suppose $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d) realizations of a random vector $X \in \mathbb{R}^p$ with Lebesgue density f. Consider the problem of estimating the quadratic functional

$$\Gamma(f) := \int_{\mathbb{R}^p} f^2(x)dx = \mathbb{E}\left[f^2(X)\right].$$

**Note.* This is a working paper and the current draft is certainly not complete. The authors accept full responsibilities for any errors in this incomplete and unpublished manuscript.

†Department of Statistics, University of Pennsylvania. Email: abhich@wharton.upenn.edu

‡Department of Statistics, University of Pennsylvania. Email: arunku@wharton.upenn.edu

A natural estimator for this functional is given by

$$\hat{\Gamma}(f) := \frac{1}{n(n-1)h_n^p} \sum_{1 \le i \ne j \le n} K\left(\frac{X_i - X_j}{h_n}\right) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}^{(-i)}(X_i), \tag{2}$$

where h_n represents the bandwidth and $\hat{f}^{(-i)}(\cdot)$ represents the leave-one-out kernel density estimator:

$$\hat{f}^{(-i)}(x) = \frac{1}{(n-1)h_n^p} \sum_{j=1, j\ne i}^{n} K\left(\frac{X_j - x}{h_n}\right)$$

Here the function $K(\cdot)$ is assumed to be symmetric and satisfies $\int_{\mathbb{R}^p} K(x)dx = 1$. This estimator was introduced by Hall and Marron (1987) and was studied thoroughly (in terms of adaptivity) for $p = 1$ in Giné and Nickl (2008).

Similarly, to estimate integrals involving the conditional expectation function from i.i.d realizations $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y), the following U-statistics appears:

$$U_n^\star := \frac{1}{n(n-1)h_n^p} \sum_{1 \le i \ne j \le n} Y_i K\left(\frac{X_i - X_j}{h_n}\right) Y_j.$$

Both the U-statistics $\hat{\Gamma}(f)$ and $U_n^\star$ are of type (1).

Apart from nonparametric statistics literature, U-statistics of type (1) also araise in relation to Hanson-Wright-type inequalities. The classical Hanson-Wright inequality concerns tail bounds for the quadratic form $G^\top AG$ where G is a standard multivariate normal random vector in $\mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ is a positive semi-definite matrix; see Theorem 3.1.9 of Giné and Nickl (2016). For further applications of Hanson-Wright inequalities, see Rudelson and Vershynin (2013) and Spokoiny and Zhilova (2013). Note that for any random vector $Y \in \mathbb{R}^n$ and matrix $A \in \mathbb{R}^{n\times n}$

$$Y^\top AY = \sum_{1 \le i,j \le n} Y_i A(i,j) Y_j,$$

where $A(i,j)$ represents the i-th row, j-th column entry in the matrix A. Thus, the quadratic form $Y^\top AY$ is also a U-statistics of type (1). Observe that in this case $w_{i,j}(\cdot,\cdot) = A(i,j)$ identically and X_i's can take arbitrary values.

Motivated by the examples above, we study the properties of the U-statistic U_n. Before proceeding further, we briefly discuss degenerate and non-degenerate U-statistics. See Serfling (1980, Chapter 5) for more details. For any sequence of functions (called kernels) $f_{i,j}(\cdot,\cdot)$ and independent random variables $Z_1, \ldots, Z_n$, a U-statistic is given by

$$T_n := \sum_{1 \le i \ne j \le n} f_{i,j}(Z_i, Z_j).$$

Note that the diagonal terms ($i = j$ cases) are ignored in the summation above. If these diagonal terms are included then the resulting statistic is called a V-statistic. The U-statistic U_n is called degenerate or canonical if the kernel functions satisfy

$$\mathbb{E}\left[f_{i,j}(Z_i, Z_j)\big|Z_i\right] = \mathbb{E}\left[f_{i,j}(Z_i, Z_j)\big|Z_j\right] = 0, \quad \text{for all} \quad 1 \le i \ne j \le n. \tag{3}$$

If the kernel functions do not satisfy (3), then the corresponding U-statistic is called non-degenerate. It is not difficult to see that a non-degenerate U-statistics can be written as a sums of independent mean zero random variables and a degenerate U-statistics:

$$T_n = \sum_{1\le i\neq j\le n} f_{i,j}^D(Z_i, Z_j) + \sum_{j=1}^{n} g_j(Z_j) + \sum_{i=1}^{n} h_i(Z_i), \tag{4}$$

where

$$\begin{aligned}
f_{i,j}^D(Z_i, Z_j) &:= f_{i,j}(Z_i, Z_j) - \mathbb{E}\left[f_{i,j}(Z_i, Z_j)\middle|Z_j\right] - \mathbb{E}\left[f_{i,j}(Z_i, Z_j)\middle|Z_i\right] + \mathbb{E}\left[f_{i,j}(Z_i, Z_j)\right], \\
g_j(Z_j) &:= \sum_{i=1, i\neq j}^{n} \left\{\mathbb{E}\left[f_{i,j}(Z_i, Z_j)\middle|Z_j\right] - \mathbb{E}\left[f_{i,j}(Z_i, Z_j)\right]\right\}, \\
h_i(Z_i) &:= \sum_{j=1, j\neq i}^{n} \left\{\mathbb{E}\left[f_{i,j}(Z_i, Z_j)\middle|Z_i\right] - \mathbb{E}\left[f_{i,j}(Z_i, Z_j)\right]\right\}.
\end{aligned} \tag{5}$$

It is clear from these expressions that the kernels $f_{i,j}^D(\cdot,\cdot)$ satisfy (3) and so are degenerate kernels. Since $T_n^{(1)}$ and $T_n^{(2)}$ in (4) are sums of independent random variables with mean zero, they can be understood easily from the classical results like the central limit theorem (asymptotically) and Bernstein/Hoeffding inequalities (non-asumptotically). For this reason, we focus mostly on the degerate part of (4) in the rest of the paper and derive non-asymptotic moment as well as tail bounds when the non-degenerate U-statistics is of the form (1). Our main tool is the decoupling inequality proved in de la Peña (1992). We refer to de la Peña and Giné (1999, Chapter 3) for more details regarding decoupling in U-statistics. After deriving non-asymptotic tail bounds for degenerate U-statistics we provide the same for supremum of degenerate U-statistics over a function class. Suppose $\mathcal{F}_n$ is a class of sequence of functions (degenerate kernels) of type $f := \{f_{i,j}^D(\cdot,\cdot) : 1 \le i \neq j \le n\}$ and define

$$\mathcal{U}_n(f) := \sum_{1\le i\neq j\le n} f_{i,j}^D(Z_i, Z_j).$$

Then $\{\mathcal{U}_n(f) : f \in \mathcal{F}_n\}$ can be viewed as a process called the U-process and we provide exponential tail bounds for the supremum:

$$\mathcal{U}_n(\mathcal{F}) := \sup_{f\in\mathcal{F}_n} |\mathcal{U}_n(f)|.$$

An important application would be the study of uniform-in-bandwidth properties of the estimator $\hat{\Gamma}(f)$ in (2), that is,

$$\sup_{h_n\in[a_n,b_n]} \left|\hat{\Gamma}(f; h_n) - \mathbb{E}\left[\hat{\Gamma}(f; h_n)\right]\right|,$$

for some numbers $a_n, b_n \in (0,1)$. Further applications can be found in de la Peña and Giné (1999, Section 5.5) and Major (2013). As a final note, we mention that even though our techniques extend to U-statistics/processes of higher order, we restrict ourselves to second order U-statistics/processes.

1.1 Literature Review

In this section, we review some of the by-now classical exponential tail bounds for degenerate U-statistics and supremum of U-processes. Proposition 2.3 of Arcones and Giné (1993) proved a Bernstein type inequality for degenerate U-statistics/processes. For the degenerate U-statistics

$$U_n := n^{-1} \sum_{1 \le i \ne j \le n} f(Z_i, Z_j),$$

with i.i.d random variables $Z_1, \ldots, Z_n$, $\sigma^2 := \mathbb{E}f^2(Z_i, Z_j)$ and $\|f\|_\infty \le C$, there exists constants $c_1, c_2 > 0$ such that for any $t > 0$,

$$\mathbb{P}\left(|U_n| \ge t\right) \le c_1 \exp\left(-\frac{c_1 t}{\sigma + (Ct^{1/2}n^{-1/2})^{2/3}}\right).$$

This tail bounds has two regimes: exponential and Weibull of order $2/3$. Because of the appearance of the variance, this tail bound provides the current rate of convergence. Theorem 3.3 of Giné et al. (2000) improved the tail bound by providing the optimal four regimes of the tail: Gaussian, exponential, Weibull of orders $2/3$ and $1/2$. Houdré and Reynaud-Bouret (2003) gave an alternative proof to the result of Giné et al. (2000) using martingale inequalities with explicit constants. In particular, Theorem 3.3 of Giné et al. (2000) shows that for all $t \ge 0$,

$$\mathbb{P}\left(|nU_n| \ge t\right) \le L \exp\left(-\frac{1}{L}\min\left\{\frac{t^2}{C^2}, \frac{t}{D}, \frac{t^{2/3}}{B^{2/3}}, \frac{t^{1/2}}{A^{1/2}}\right\}\right),$$

for some constants A, B, C, D and L. The main disadvantage of the results above is the restrictive boundedness assumption. Theorem 3.2 of Giné et al. (2000) actually applies without the boundedness condition but the tail bound thus obtained is sub-optimal. For example if $f(Z_i, Z_j) = Y_i g(X_i, X_j) Y_j$ where $Z_i = (X_i, Y_i)$, $\|g\|_\infty \le C < \infty$ and Y_i's are mean zero sub-Weibull variables of order $\alpha > 0$, that is, $\mathbb{P}(|Y_i| \ge t|X_i) \le 2\exp(-t^\alpha)$. Then Theorem 3.2 of Giné et al. (2000) implies a tail bound of the form:

$$\mathbb{P}\left(|nU_n| \ge t\right) \le L \exp\left(-\frac{1}{L}\min\left\{\frac{t^2}{C^2}, \frac{t}{D}, \frac{t^{\alpha_1}}{B^{\alpha_1}}, \frac{t^{\alpha_2}}{A^{\alpha_2}}\right\}\right),$$

where $\alpha_1^{-1} = (3/2 + 1/\alpha)$ and $\alpha_2^{-1} = (2 + 2/\alpha)^{-1}$. This is sub-optimal in comparison with the results of Kolesko and Latała (2015, Example 3). On the other hand, the results of Kolesko and Latała (2015) do not get the correct rate of convergence as can be obtained from the results of Giné et al. (2000). This is because the bound of Kolesko and Latała (2015) does not depend on the variance. We are not aware of any tail bounds in the literature that implies the correct rate of convergence as well as the optimal tail behavior.

In regards to the tail bounds for U-processes, some of the important works are Adamczak (2006), Clémençon et al. (2008) and Major (2013). The latter two papers only consider bounded kernels and the bounds of Adamczak (2006) are written in terms of functionals that are in general hard to control. The results of Major (2005) and Major (2013) apply only to bounded kernels and are written for VC class $\mathcal{F}_n$ but imply the correct rate of convergence. However, the results there do not show the optimal four regimes in the tail behavior. Theoerem 11 of Clémençon et al. (2008) is written as a deviation inequality but does not imply the correct rate of convergence. For instance, if $f(X_i, X_j) = \varepsilon_i \varepsilon_j K((X_i - X_j)/h)$ with ε_i being Rademacher random variables indepenedent of $X_i \in \mathbb{R}^p$, then the rate of convergence of

$$T_n := \sup_{h \in \{h_n\}} \left| \sum_{1 \le i \neq j \le n} \varepsilon_i \varepsilon_j K\left(\frac{X_i - X_j}{h}\right) \right|,$$

from Theorem 11 of Clémençon et al. (2008) is $n \|K\|_\infty = O(n)$ (because of Fn in the moment bound) but the correct rate of convergence is $nh_n^{p/2}$ (that can be obtained by calculating the variance). As in the case of U-statistics, we are not aware of any tail bound results that can obtain the correct rate of convergence and apply to unbounded kernels. Using the techniques of truncation, decoupling technique and the entropy method of Boucheron et al. (2005), we prove a deviation inequality for degenerate U-processes that implies the correct rate of convergence and the optimal tail behavior.

The remaining article is organized as follows. In Section 2 we prove exponential tail bounds for degenerate U-statistics of the product form kernel in (1). In Section 3 we prove a deviation bound for degenerate U-processes and also provide maximal inequalities to control the expectation of the maximum.

2 Tail Bounds for Degenerate U–Statistics

Recall the setting of the random variables and U-statistics from Section 1. There are n independent random variables $Z_1 = (X_1, Y_1), Z_2 = (X_2, Y_2), \ldots, Z_n = (X_n, Y_n)$ on some measurable space and sequence of functions $\{w_{i,j}(\cdot, \cdot) : 1 \le i \neq j \le n\}$. Consider, for functions $\phi(\cdot)$ and $\psi(\cdot)$ the U-statistic

$$U_n := \sum_{1 \le i \neq j \le n} f_{i,j}(Z_i, Z_j), \quad \text{where} \quad f_{i,j}(Z_i, Z_j) := \phi(Z_i) w_{i,j}(X_i, X_j) \psi(Z_j). \tag{6}$$

The kernels $f_{i,j}(\cdot, \cdot)$ are not required to be degenerate here. We will derive moment and tail bounds for the degenerate version of the U-statistics T_n^D given by

$$U_n^D := \sum_{1 \le i \neq j \le n} f_{i,j}^D(Z_i, Z_j),$$

for the kernel $f_{i,j}^D(\cdot, \cdot)$ defined in (5). We first prove a basic lemma that reduces the problem of moment bounds on U_n^D to a symmetrized version of U_n; see Theorem 3.5.3 of de la Peña and Giné (1999). For any random variable W, set $\|W\|_p = (\mathbb{E}[|W|^p])^{1/p}$ for $p \ge 1$.

Lemma 1. *For any $p \ge 1$,*

$$\left\|U_n^D\right\|_p \le C \left\| \sum_{1 \le i \neq j \le n} \varepsilon_i \varepsilon_j' f_{i,j}(Z_i, Z_j') \right\|_p,$$

for Rademacher random variables $(\varepsilon_i, \varepsilon_i' : 1 \le i \le n)$. Here C can be taken to be 192 and $Z_1' = (X_1', Y_1'), \ldots, Z_n' = (X_n', Y_n')$ represents an independent of n independent random variables such that Z_i is identically distributed as Z_i for $1 \le i \le n$.

The proof of this lemma (in Section A) is based on the by-now classical decoupling inequalities of de la Peña (1992) and de la Peña and Giné (1999, Chapter 3). The result also holds in case of degenerate U-processes and does not require the special structure of the kernels $f_{i,j}(\cdot, \cdot)$ in (6).

To prove moment and tail bounds for degenerate second order U-statistics with unbounded kernels, we use the following assumptions. Define a random variable W to be sub-Weibull of order $\alpha > 0$ if $\|W\|_{\psi_\alpha} < \infty$, where $\psi_\alpha(x) = \exp(x^\alpha) - 1$ for $x \ge 0$ and

$$\|W\|_{\psi_\alpha} = \inf\left\{C \ge 0 : \mathbb{E}\left[\psi_\alpha\left(|W|/C\right)\right] \le 1\right\}.$$

Several properties of sum of independent sub-Weibull random variables are derived in Kuchibhotla and Chakrabortty (2018). The main focus of this section is to extend these results to degenerate U-statistics of special type. Consider the following assumptions.

(A1) There exists constants $0 < \alpha, \beta, C_\phi, C_\psi < \infty$ such that

$$\max_{1 \le i \le n} \mathbb{E}\left[\exp\left(\frac{|\phi(Z_i)|^\alpha}{C_\phi^\alpha}\right) \big| X_i\right] \le 2, \quad \text{and} \quad \max_{1 \le i \le n} \mathbb{E}\left[\exp\left(\frac{|\psi(Z_i)|^\beta}{C_\psi^\beta}\right) \big| X_i\right] \le 2,$$

hold almost surely.

(A2) The functions $\{w_{i,j}(\cdot, \cdot) : 1 \le i \neq j \le n\}$ are all uniformly bounded, that is,

$$\max_{1 \le i \neq j \le n} \sup_{(x,x') \in \mathfrak{X} \times \mathfrak{X}} \left|w_{i,j}(x, x')\right| \le B_w.$$

The main technique in our proof is truncation and Hoffmann-Jørgensen's inequality. Assumption **(A1)** implies that conditional on X_i's the maximum of $\phi(Y_i)$ is at most a polynomial of $\log n$ (in rate). This along with Assumption **(A2)** allows us to apply truncation at this rate and study the truncated part using the sharp results of Giné et al. (2000). The unbounded parts of smaller order are controlled using Hoffmann-Jørgensen's inequality. The bound B_w in Assumption **(A2)** is allowed to grown in n and all the kernels are also allowed to be function of n. All the results to be presented here are non-asymptotic. For more applications of this technique see Kuchibhotla and Chakrabortty (2018).

Define

$$T_\phi := 8\mathbb{E}\left[\max_{1\le i\le n}|\phi(Z_i)|\,\middle|X_1,\ldots,X_n\right],\quad T_\psi := 8\mathbb{E}\left[\max_{1\le i\le n}|\psi(Z_i)|\,\middle|X_1,\ldots,X_n\right],$$

and the truncated random variables

$$\begin{aligned} \Phi_{i,1} &:= \phi(Z_i)\mathbb{1}\{|\phi(Z_i)|\le T_\phi\}, \quad\text{and}\quad \Phi_{i,2} := \phi(Z_i)\mathbb{1}\{|\phi(Z_i)| > T_\phi\},\\ \Psi'_{j,1} &:= \psi(Z'_j)\mathbb{1}\{|\psi(Z'_j)|\le T_\psi\}, \quad\text{and}\quad \Psi'_{j,2} := \psi(Z'_j)\mathbb{1}\{|\psi(Z'_j)| > T_\psi\}. \end{aligned} \tag{7}$$

It is clear that $\phi(Z_i) = \Phi_{i,1} + \Phi_{i,2}$ and $\psi(Z'_j) = \Psi'_{j,1} + \Psi'_{j,2}$. Based on these, note that

$$\begin{aligned} \phi(Z_i)w_{i,j}(X_i,X'_j)\psi(Z'_j) &= \Phi_{i,1}w_{i,j}(X_i,X_j)\Psi'_{j,1} + \Phi_{i,2}w_{i,j}(X_i,X_j)\Psi'_{j,1}\\ &\quad + \Phi_{i,1}w_{i,j}(X_i,X_j)\Psi'_{j,2} + \Phi_{i,2}w_{i,j}(X_i,X_j)\Psi'_{j,2}. \end{aligned} \tag{8}$$

The first term on the right hand side is bounded by $T_\phi B_w T_\psi$. The second and third terms are non-zero only when $\Phi_{i,2}$ and $\Psi'_{j,2}$, are respectively non-zero, which can only happen with only a small probability under Assumption **(A1)**. Finally, the fourth term can be non-zero only if both $\Phi_{i,2}$ and $\Psi'_{j,2}$ are non-zero which can happend with even smaller probability. These four terms leads to four different degenerate U-statistics that will be controlled separately in Section A.1 to prove the following result. We need the following notation: for $1\le i,j\le n$,

$$\sigma^2_{i,\phi}(x) = \mathbb{E}[\phi^2(Z_i)\big|X_i = x] \quad\text{and}\quad \sigma^2_{j,\psi}(x) = \mathbb{E}[\psi^2(Z_j)\big|X_j = x].$$

Define $\Lambda_2 := C_\phi C_\psi B_w(\log n)^{\alpha^{-1}+\beta^{-1}}$ and

$$\begin{aligned} \Lambda_{1/2} &:= \left(\sum_{1\le i\ne j\le n}\mathbb{E}\left[\sigma^2_{i,\phi}(X_i)w^2_{i,j}(X_i,X_j)\sigma^2_{j,\psi}(X_j)\right]\right)^{1/2},\\ \Lambda_1 &:= \sup\Bigg\{\sum_{1\le i\ne j\le n}\mathbb{E}\left[q_i(X_i)\sigma_{i,\phi}(X_i)w_{i,j}(X_i,X_j)\sigma_{j,\psi}(X_j)p_j(X_j)\right]:\\ &\qquad\qquad \sum_{j=1}^n\mathbb{E}\left[q_i^2(X_i)\right]\le 1,\ \sum_{i=1}^n\mathbb{E}\left[p_j^2(X_j)\right]\le 1\Bigg\},\\ \Lambda^{(\alpha)}_{3/2} &:= C_\phi(\log n)^{1/\alpha}\sup_x\max_{1\le i\le n}\left(\sum_{j=1}^n\mathbb{E}\left[w^2_{i,j}(x,X_j)\sigma^2_{j,\psi}(X_j)\right]\right)^{1/2}\\ \Lambda^{(\beta)}_{3/2} &:= C_\psi(\log n)^{1/\beta}\sup_x\max_{1\le j\le n}\left(\sum_{i=1}^n\mathbb{E}\left[w^2_{i,j}(X_i,x)\sigma^2_{i,\phi}(X_i)\right]\right)^{1/2}\\ \Lambda_{\alpha^*} &:= (\log n)^{1/2}\Lambda^{(\alpha)}_{3/2} + (\log n)\Lambda_2,\\ \Lambda_{\beta^*} &:= (\log n)^{1/2}\Lambda^{(\beta)}_{3/2} + (\log n)\Lambda_2. \end{aligned}$$

The quantities $\Lambda_{1/2}, \Lambda_1, \Lambda_{3/2}^{(\alpha)}, \Lambda_{3/2}^{(\beta)}, \Lambda_2$ also appear in the case of bounded kernels as shown in Theorem 3.2 of Giné et al. (2000).

Theorem 1. *Under Assumptions **(A1)** and **(A2)**, there exists constant $K > 0$ (depending only on α, β) such that for all $p \geq 1$*

$$\begin{aligned}\left\|U_n^D\right\|_p \leq Kp^{1/2}\Lambda_{1/2} + Kp\Lambda_1 + Kp^{1/\alpha^*}\Lambda_{\alpha^*} + Kp^{1/\beta^*}\Lambda_{\beta^*} \\ + Kp^{1/2+1/\alpha^*}\Lambda_{3/2}^{(\alpha)} + Kp^{1/2+1/\beta^*}\Lambda_{3/2}^{(\beta)} + Kp^{1/\alpha^*+1/\beta^*}\Lambda_2.\end{aligned}$$

Here $\alpha^* := \min\{\alpha, 1\}$ and $\beta^* := \min\{\beta, 1\}$. By Markov's inequality, there exists a constant $K' > 0$ such that for any $t \geq 0$,

$$\mathbb{P}\left(|U_n^D| \geq K'\mathcal{T}_{\alpha,\beta}(t)\right) \leq 2\exp(-t), \tag{9}$$

where

$$\begin{aligned}\mathcal{T}_{\alpha,\beta}(t) := \sqrt{t}\Lambda_{1/2} + t\Lambda_1 + t^{1/\alpha^*}\Lambda_{\alpha^*} + t^{1/\beta^*}\Lambda_{\beta^*} \\ + t^{1/2+1/\alpha^*}\Lambda_{3/2}^{(\alpha)} + t^{1/2+\beta^*}\Lambda_{3/2}^{(\beta)} + t^{1/\alpha^*+1/\beta^*}\Lambda_2.\end{aligned}$$

Remark 2.1 (Comparison with previous results) As noted in the introduction, an important feature of our result is that the kernel is allowed to be unbounded with proper tail behavior. The tail of the degenerate U-statistics as shown in (9) has *seven* different regimes, the prominent ones being the Gaussian and exponential parts. These seven regimes collapse to *five* if $\alpha = \beta$. In particular, if $\alpha = \beta \leq 1$, then for $p \geq 1$,

$$\begin{aligned}\left\|U_n^D\right\|_p \leq K\mathbf{p}^{\mathbf{1/2}}\Lambda_{1/2} + K\mathbf{p}\Lambda_1 + K\mathbf{p}^{\mathbf{1}/\alpha}\left[(\log n)^{1/2}\left\{\Lambda_{3/2}^{(\alpha)} + \Lambda_{3/2}^{(\beta)}\right\} + (\log n)\Lambda_2\right] \\ + K\mathbf{p}^{\mathbf{1/2+1}/\alpha}\left[\Lambda_{3/2}^{(\alpha)} + \Lambda_{3/2}^{(\beta)}\right] + K\mathbf{p}^{\mathbf{1}/\alpha+\mathbf{1}/\beta}\Lambda_2.\end{aligned}$$

If $\alpha = \beta = \infty$, then our assumption **(A1)** implies boundedness of the kernels. In this case, only four regimes remain and these four regimes coincide with those shown in Theorem 3.2 of Giné et al. (2000). Additionally in the case of bounded kernels ($\alpha = \beta = \infty$), our result essentially coincides with Theorem 3.2 of Giné et al. (2000) except for the additional $\sqrt{\log n}$ and $\log n$ factors. We believe these to be artifacts of our proof although we do not know if these factors can be avoided in the general setting. It is worth mentioning here that we could not use the proof technique of Giné et al. (2000) to avoid these factors. ◇

3 Tail Bounds for Degenerate U–Processes

In this section, we generalize Theorem 1 to degenerate U-processes. Consider

$$\mathcal{U}_n(\mathcal{W}) := \sup_{w\in\mathcal{W}} |\mathcal{U}_n(w)|, \quad \text{where} \quad \mathcal{U}_n(w) := \sum_{1\leq i\neq j\leq n} \varepsilon_i\phi(Z_i)w_{i,j}(X_i, X_j')\psi(Z_j')\varepsilon_j',$$

for some function class $\mathcal{W}$ with elements of the type $w = (w_{i,j})_{1\le i\neq j\le n}$. If $\mathcal{W}$ is a singleton, then this reduces to the U-statistic studied in Section 2. Here $\varepsilon_1,\ldots,\varepsilon_n$ denote an independent sequence of Rademacher random variables as before. For simplicity, we consider the symmetrized version and by Lemma 1 the results also hold for the original degenerate U-process; see Theorem 3.5.3 of de la Peña and Giné (1999) for details.

U-processes were introduced in Nolan and Pollard (1987) to study cross-validation in the context of kernel density estimation. They studied uniform almost sure limit theorems and established the rate of convergence. These results parallel the Glivenko-Cantelli theorems well-known for empirical processes. Functional limit theorems were established in Nolan et al. (1988). Exponential tail bounds that parallel the classical Bernstein's inequality for non-degenerate and degenerate U-statistics were given in Arcones and Giné (1993). They also established LLN and CLT type results under various metric entropy conditions. Most of these results require boundedness of the kernel functions. Being asymptotic in nature, some of these results can be extended to the case of unbounded kernels using a truncation argument. Finite sample concentration inequalities for degenerate unbounded U-processes are not readily available.

The only work (we are aware of) that provides general results for U-processes applicable to $\mathcal{U}_n(\mathcal{W})$ is Adamczak (2006). In this work, degenerate U-processes of arbitrary order were considered. However, the moment bounds for U-processes in this work depend further on the moment bounds of some complicated degenerate U-processes of lower order. Furthermore, the tail behavior thus obtained is not sharp for unbounded U-processes.

To avoid measurability issues for $\mathcal{U}_n(\mathcal{W})$, we use either of the following conventions. One simple assumption on $\mathcal{W}$ used in van der Vaart and Wellner (1996) that implies measurability is separability and in this case we can take $\mathcal{W}$ to be a dense countable subset of $\mathcal{W}$. Another convention used in Talagrand (2014) is to define for any $\mathcal{W}$ and increasing function $f(\cdot)$,

$$\mathbb{E}\left[f\left(\mathcal{U}_n(\mathcal{W})\right)\right] := \sup\left\{\mathbb{E}\left[f(\mathcal{U}_n(\mathcal{F}))\right] :\, \mathcal{F}\subseteq\mathcal{W} \text{ a finite subset}\right\}.$$

Based on either convention, we treat $\mathcal{W}$ as a countable set for the remaining part of this section.

One "simple" way to obtain tail bounds for $\mathcal{U}_n(\mathcal{W})$ is via generic chaining as follows: First apply Theorem 1 for $\mathcal{U}_n(w) - \mathcal{U}_n(w')$ for functions $w, w' \in \mathcal{W}$. The tail bound (9) provides a mixed tail in terms of various semi-metrics on $\mathcal{W}$. Using these and following the proof of classical generic chaining bound (e.g., Theorem 3.5 of Dirksen (2015)), one can obtain tail bounds for U-processes in terms of γ-functionals; see Talagrand (2014) and Dirksen (2015) for details. A problem with this approach is the complication in controlling the γ-functionals. This approach with Dudley's chaining (instead of generic chaining) was used for bounded kernel U-processes in Nolan and Pollard (1987) and Nolan et al. (1988).

In the following, we first provide a deviation inequality for $\mathcal{U}_n(\mathcal{W})$ and then prove a maximal inequality to control the expectations appearing in the deviation inequality. For these results, we consider the following generalization of assumption **(A2)**.

(A2′) The functions $\{w : w \in \mathcal{W}\}$ are all uniformly bounded, that is,

$$\sup_{w\in\mathcal{W}} \sup_{(x,x')\in\mathfrak{X}\times\mathfrak{X}} \max_{1\le i\neq j\le n} \left|w_{i,j}(x,x')\right| \le B_{\mathcal{W}}.$$

We will use the notation of $\Phi_{i,1}, \Phi_{i,2}, \Psi'_{i,1}, \Psi'_{i,2}$ given in (7). For the main result of this section, define

$$\begin{aligned}
\Lambda_2(\mathcal{W}) &:= (\log n)^{\alpha^{-1}+\beta^{-1}} C_\phi C_\psi B_{\mathcal{W}},\\
E_{n,1}(\mathcal{W}) &:= C_\psi(\log n)^{1/\beta} \sup_{x\in\mathfrak{X}} \max_{1\le j\le n} \mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\sum_{i=1,i\neq j}^{n} \varepsilon_i\Phi_{i,1}w_{i,j}(X_i,x)\right|\right],\\
E_{n,2}(\mathcal{W}) &:= C_\phi(\log n)^{1/\alpha} \sup_{x\in\mathfrak{X}} \max_{1\le i\le n} \mathbb{E}\left[\sup_{w\in\mathcal{W}}\left|\sum_{j=1,j\neq i}^{n} \varepsilon'_j\Psi'_{j,1}w_{i,j}(x,X'_j)\right|\right],\\
\mathfrak{W}_{n,1}(\mathcal{W}) &:= \mathbb{E}\left[\sup_{w\in\mathcal{W}}\sup_{\{p_j\}}\sum_{1\le i\neq j\le n}\varepsilon_i\Phi_{i,1}\int p_j(x)\sigma_{j,\psi}(x)w_{i,j}(X_i,x)P_{X_j}(dx)\right],\\
\mathfrak{W}_{n,2}(\mathcal{W}) &:= \mathbb{E}\left[\sup_{w\in\mathcal{W}}\sup_{\{q_i\}}\sum_{1\le i\neq j\le n}\varepsilon'_j\Psi'_{j,1}\int q_i(x)\sigma_{i,\phi}(x)w_{i,j}(x,X'_j)P_{X_i}(dx)\right],\\
\Sigma^{1/2}_{n,1}(\mathcal{W}) &:= C_\psi(\log n)^{1/\beta}\sup_{x\in\mathfrak{X}}\sup_{w\in\mathcal{W}}\max_{1\le j\le n}\left(\sum_{i=1,i\neq j}^{n}\mathbb{E}[\sigma^2_{i,\phi}(X_i)w^2_{i,j}(X_i,x)]\right)^{1/2},\\
\Sigma^{1/2}_{n,2}(\mathcal{W}) &:= C_\phi(\log n)^{1/\alpha}\sup_{x\in\mathfrak{X}}\sup_{w\in\mathcal{W}}\max_{1\le i\le n}\left(\sum_{j=1,j\neq i}^{n}\mathbb{E}\left[\sigma^2_{j,\psi}(X_j)w^2_{i,j}(x,X_j)\right]\right)^{1/2},\\
\|(\phi w\psi)_{\mathcal{W}}\|_{2\to2} &:= \sup_{w\in\mathcal{W}}\sup_{\{q_i\}}\sup_{\{p_j\}}\sum_{1\le i\neq j\le n}\mathbb{E}\left[q_i(X_i)\sigma_{i,\phi}(X_i)w_{i,j}(X_i,X'_j)\sigma_{j,\psi}(X'_j)p_j(X'_j)\right].
\end{aligned}$$

Here in the definitions, the supremum over $\{q_i\}$ (or $\{p_j\}$) represents supremum over all function $(q_1,\ldots,q_n)$ (or $(p_1,\ldots,p_n)$)satisfying

$$\sum_{i=1}^{n}\int q_i^2(x)P_{X_i}(dx)\le 1,\quad\text{and}\quad \sum_{j=1}^{n}\int p_j^2(x)P_{X_i}(dx)\le 1,$$

where $P_{X_i}(\cdot)$ denotes the probability measure of X_i. Note that $\|(\phi w\psi)_{\mathcal{W}}\|_{2\to2}$ is similar to Λ_1 defined for Theorem 1.

Theorem 2. *Under assumptions (A1) and (A2′), there exists a constant $K>0$ (de-*

pending only on α, β) such that for all $p \geq 1$

$$\begin{aligned}\|\mathcal{U}_n(\mathcal{W})\|_p \leq K\mathbb{E}\left[\mathcal{U}_n^{(1)}(\mathcal{W})\right] &+ Kp^{1/2}(\mathfrak{W}_{n,1}(\mathcal{W}) + \mathfrak{W}_{n,2}(\mathcal{W})) + Kp\,\|(\phi w \psi)_{\mathcal{W}}\|_{2\to 2} \\ &+ Kp^{1/\alpha^*}\left[E_{n,2}(\mathcal{W}) + \Sigma_{n,2}^{1/2}(\mathcal{W})\sqrt{\log n} + \Lambda_2(\mathcal{W})\log n\right] \\ &+ Kp^{1/\beta^*}\left[E_{n,1}(\mathcal{W}) + \Sigma_{n,1}^{1/2}(\mathcal{W})\sqrt{\log n} + \Lambda_2(\mathcal{W})\log n\right] \\ &+ Kp^{1/2+1/\alpha^*}\Sigma_{n,2}^{1/2}(\mathcal{W}) + Kp^{1/2+1/\beta^*}\Sigma_{n,1}^{1/2}(\mathcal{W}) + Kp^{1/\alpha^*+1/\beta^*}\Lambda_2(\mathcal{W}).\end{aligned}$$

Proof. See Appendix B.1 for a proof. □

If $\mathcal{W}$ is a singleton set, then the above result reduces to Theorem 1. From the moment bound above, it is easy to derive a tail bound using Markov's inequality. In comparison, we again get *seven* different tail regimes that again reduce to *five* if $\alpha = \beta$. Unlike the result of Adamczak (2006), the moment bound in Theorem 2 only depends on some expectations. An additional advantage of Theorem 2 is that all the expectations only involve bounded random variables.

3.1 Maximal Inequality for Bounded Degenerate U-Processes

To apply Theorem 2, we need to control various expectations appearing on the right hand side of the moment bound there. Expect for $\mathbb{E}[\mathcal{U}_n^{(1)}(\mathcal{W})]$, all the other quantities are maximal inequalities related to empirical processes. See van der Vaart and Wellner (2011) and Lemmas 3.4.2-3.4.3 of van der Vaart and Wellner (1996) for maximal inequalities of empirical processes. In this section, we provide a maximal inequality for $\mathcal{U}_n^{(1)}(\mathcal{W})$. For independent and identically distributed random variables, Chen and Kato (2017, Theorem 5.1) provide a maximal inequality for degenerate U-processes of arbitrary order. This result is similar to Theorem 2.1 of van der Vaart and Wellner (2011) for empirical processes. The same proof as in Chen and Kato (2017) does not provide the "correct" bound in the case of possibly non-identically distributed observations since they use Hoeffding averging which can lead to sub-optimal rate if the observations are not identically distributed. A modification of the proof leads to the maximal inequality below.

For any $\eta > 0$, function class $\mathcal{F}$ containing functions $f = (f_{i,j})_{1\leq i\neq j\leq n} : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ and a discrete probability measure Q with support $\{z_1, \ldots, z_t\}$, let $N(\eta, \mathcal{F}, \|\cdot\|_{2,Q})$ denotes the minimum m such that there exists $f^{(1)}, f^{(2)}, \ldots, f^{(m)} \in \mathcal{F}$ satisfying

$$\sup_{f\in\mathcal{F}} \inf_{1\leq j\leq m} \left\|f - f^{(j)}\right\|_{2,Q} \leq \eta,$$

where for $f \in \mathcal{F}$,

$$\|f\|_{2,Q}^2 := \frac{\sum_{1\leq i\neq j\leq t} f_{i,j}^2(z_i, z_j)Q(\{z_i\})Q(\{z_j\})}{\sum_{1\leq i\neq j\leq t} Q(\{z_i\})Q(\{z_j\})}.$$

Note that the right hand side is expectation with respect to the measure induced on $\{(z_i, z_j) : 1 \leq i \neq j \leq t\}$. Define the uniform entropy integral needed for U-processes is given by

$$J_2(\delta, \mathcal{F}, \|\cdot\|_2) := \sup_Q \int_0^\delta \log N(\eta \|F\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}) d\eta.$$

Here $F = (F_{i,j})_{1 \leq i \neq j \leq n}$ represents the envelope function for $\mathcal{F}$ satisfying $|f_{i,j}(x, x')| \leq F_{i,j}(x, x')$ for all $f \in \mathcal{F}, x, x' \in \mathcal{X}$ and the supremum is taken over all discrete probability measures Q supported on $\mathcal{X} \times \mathcal{X}$. The following Lemma proves a maximal inequality using Theorem 5.1.4 of de la Peña and Giné (1999). The proof is very similar to that of Theorem 5.1 of Chen and Kato (2017) which itself was based on the proof of Theorem 2.1 of van der Vaart and Wellner (2011).

Theorem 3. *Suppose $\mathcal{F}$ represent a class of real-valued functions $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ uniformly bounded by R with the envelope function F. Then there exists a universal constant $C > 0$ such that*

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} \left| \frac{\sum_{1 \leq i \neq j \leq n} \epsilon_i \epsilon_j f_{i,j}(X_i, X_j)}{\sqrt{n(n-1)}} \right| \right] \leq C \|F\|_{2,P} J_2(a, \mathcal{F}, \|\cdot\|_2) \left[1 + \frac{J_2(a, \mathcal{F}, \|\cdot\|_2) b^2}{a^2}\right],$$

for any $a \geq A_n$ and $b \geq B_n$, where $B_n^2 = R/(n \|F\|_{2,P})$,

$$\begin{aligned}
\|F\|_{2,P}^2 &:= \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \mathbb{E}[F_{i,j}^2(X_i, X_j)], \\
A_n^2 &:= \|F\|_{2,P}^{-2} \left[\Gamma_{n,1}^2(\mathcal{F}) + \Gamma_{n,2}^2(\mathcal{F}) + \Sigma_n^2(\mathcal{F})\right], \\
\Gamma_{n,1}^2(\mathcal{F}) &:= \mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{n(n-1)} \left| \sum_{1 \leq i \neq j \leq n} \left\{\mathbb{E}\left[f_{i,j}^2(X_i, X_j) | X_i\right] - \mathbb{E}\left[f_{i,j}^2(X_i, X_j)\right]\right\} \right| \right], \\
\Gamma_{n,2}^2(\mathcal{F}) &:= \mathbb{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{n(n-1)} \left| \sum_{1 \leq i \neq j \leq n} \left\{\mathbb{E}\left[f_{i,j}^2(X_i, X_j) | X_j\right] - \mathbb{E}\left[f_{i,j}^2(X_i, X_j)\right]\right\} \right| \right], \\
\Sigma_n^2(\mathcal{F}) &:= \sup_{f \in \mathcal{F}} \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \mathbb{E}\left[f_{i,j}^2(X_i, X_j)\right].
\end{aligned}$$

APPENDIX

A Proofs of Results in Section 2

Throughout the proofs in all the appendices to follow, we use the notation

$$\mathcal{Z}_n' := \{(Z_1', \varepsilon_1'), \ldots, (Z_n', \varepsilon_n')\} \quad \text{and} \quad \mathcal{Z}_n := \{(Z_1, \varepsilon_1), \ldots, (Z_n, \varepsilon_n)\}.$$

Note that this is different from $\mathcal{Z}_n'$ and $\mathcal{Z}_n$ defined in the main text.

Proof of Lemma 1. From Theorem 3.1.1 of de la Peña and Giné (1999) and following the arguments similar to those in Theorem 3.5.3 of de la Peña and Giné (1999), we get for all $p \ge 1$

$$\left\|T_n^D\right\|_p \le 48 \left\| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j' f_{i,j}^D(Z_i, Z_j') \right\|_p,$$

where $\varepsilon_i, \varepsilon_i', 1 \le i \le n$ are Rademacher random variables independent of $(Z_i, Z_i'), 1 \le i \le n$. Note from (5) that

$$\begin{aligned}\varepsilon_i \varepsilon_j' f_{i,j}^D(Z_i, Z_j') &= \varepsilon_i \varepsilon_j' f_{i,j}(Z_i, Z_j') - \varepsilon_i \varepsilon_j' \int f_{i,j}(z, Z_j') P_i(dz) \\ &\quad - \varepsilon_i \varepsilon_j' \int f_{i,j}(Z_i, z) P_j(dz) + \varepsilon_i \varepsilon_j' \iint f_{i,j}(z, z') P_i(dz) P_j(dz).\end{aligned}$$

Here P_i represents the probability measure of Z_i for $1 \le i \le n$. By Jensen's inequality, it is clear that for $p \ge 1$,

$$\begin{aligned}\left\| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j' \int f_{i,j}(z, Z_j') P_i(dz) \right\|_p &\le \left\| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j' f_{i,j}(Z_i, Z_j') \right\|_p, \\ \left\| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j' \int f_{i,j}(Z_i, z) dP_j(z) \right\|_p &\le \left\| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j' f_{i,j}(Z_i, Z_j') \right\|_p, \\ \left\| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j' \iint f_{i,j}(z, z') P_i(dz) P_j(dz) \right\|_p &\le \left\| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j' f_{i,j}(Z_i, Z_j') \right\|_p.\end{aligned}$$

Therefore, for $p \ge 1$,

$$\left\|T_n^D\right\|_p \le 192 \left\| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j' f_{i,j}(Z_i, Z_j') \right\|_p$$

□

A.1 Proof of Theorem 1

Based on the basic decomposition (8), we get

$$\sum_{1\le i\neq j\le n} \varepsilon_i \phi(Z_i) w_{i,j}(X_i, X'_j)\psi(Z'_j)\varepsilon'_j = \mathcal{U}_n^{(1)} + \mathcal{U}_n^{(2)} + \mathcal{U}_n^{(3)} + \mathcal{U}_n^{(4)},$$

where

$$\begin{aligned}
\mathcal{U}_n^{(1)} &:= \sum_{1\le i\neq j\le n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X'_j)\Psi'_{j,1}\varepsilon'_j, \\
\mathcal{U}_n^{(2)} &:= \sum_{1\le i\neq j\le n} \varepsilon_i \Phi_{i,2} w_{i,j}(X_i, X'_j)\Psi'_{j,1}\varepsilon'_j, \\
\mathcal{U}_n^{(3)} &:= \sum_{1\le i\neq j\le n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X'_j)\Psi'_{j,2}\varepsilon'_j, \\
\mathcal{U}_n^{(4)} &:= \sum_{1\le i\neq j\le n} \varepsilon_i \Phi_{i,2} w_{i,j}(X_i, X'_j)\Psi'_{j,2}\varepsilon'_j.
\end{aligned} \tag{10}$$

It is easy to verify that $\mathcal{U}_n^{(k)}, 1 \le k \le 4$ are all degenerate U-statistics. From Theorem 3.2 of Giné et al. (2000), we get that there exists a constant $K > 0$ such that for all $p \ge 1$,

$$\left\| \mathcal{U}_n^{(1)} \right\|_p \le K \left[\sqrt{p}A + pB + p^{3/2}C + p^2 D \right],$$

where

$$\begin{aligned}
A &:= \left(\sum_{1\le i\neq j\le n} \mathbb{E}\left[\Phi_{i,1}^2 w_{i,j}^2(X_i, X'_j) \left(\Psi'_{i,1}\right)^2 \right] \right)^{1/2}, \\
B &:= \sup \left\{ \mathbb{E} \sum_{1\le i\neq j\le n} \varepsilon_i \xi_i(\varepsilon_i, Z_i)\Phi_{i,1} w_{i,j}(X_i, X'_j)\Psi'_{i,1}\zeta_j(\varepsilon'_j, Z'_j)\varepsilon'_j : \right. \\
&\qquad \left. \mathbb{E}\sum_{i=1}^n \xi_i^2(\varepsilon_i, Z_i) \le 1,\ \mathbb{E}\sum_{j=1}^n \zeta_i^2(\varepsilon'_i, Z'_i) \le 1 \right\}, \\
C^p &:= \mathbb{E}\left(\max_{1\le i\le n} \mathbb{E}\left[\sum_{j=1}^n \Phi_{i,1}^2 w_{i,j}^2(X_i, X'_j) \left(\Psi'_{i,1}\right)^2 \middle| X_i, Y_i \right] \right)^{p/2} \\
&\quad + \mathbb{E}\left(\max_{1\le j\le n} \mathbb{E}\left[\sum_{i=1}^n \Phi_{i,1}^2 w_{i,j}^2(X_i, X'_j) \left(\Psi'_{i,1}\right)^2 \middle| X'_j, Y'_j \right] \right)^{p/2} \\
D^p &:= \mathbb{E}\left[\max_{1\le i\neq j\le n} |\Phi_{i,1} w_{i,j}(X_i, X'_j)\Psi'_{j,1}|^p \right].
\end{aligned} \tag{11}$$

It is clear that

$$A^2 \leq \sum_{1\leq i\neq j\leq n} \mathbb{E}\left[\phi^2(Y_i)w_{i,j}^2(X_i, X_j)\psi^2(Y_j)\right] = \sum_{1\leq i\neq j\leq n} \mathbb{E}\left[\sigma_{i,\phi}^2(X_i)w_{i,j}^2(X_i, X_j)\sigma_{j,\psi}^2(X_j)\right].$$

The quantity B appears as the square root of the wimpy variance of the supremum of an empirical process; see Boucheron et al. (2013, page 314). Lemma 4 of Section A.2 implies that

$$B \leq \sup\left\{\sum_{1\leq i\neq j\leq n} \mathbb{E}\left[q_i(X_i)\sigma_{i,\phi}(X_i)w_{i,j}(X_i, X_j)\sigma_{j,\psi}(X_j)p_j(X_j)\right]:\right.$$
$$\left.\sum_{j=1}^{n}\mathbb{E}\left[q_i^2(X_i)\right] \leq 1,\ \sum_{i=1}^{n}\mathbb{E}\left[p_j^2(X_j)\right] \leq 1\right\}.$$

For bounding C, note that

$$\mathbb{E}\left[\sum_{j=1}^{n}\Phi_{i,1}^2 w_{i,j}^2(X_i, X_j')\left(\Psi_{j,1}'\right)^2\Big|X_i, Y_i\right] \leq T_\phi^2\sup_x\sum_{j=1}^{n}\mathbb{E}\left[w_{i,j}^2(x, X_j)\sigma_{j,\psi}^2(X_j)\right],$$
$$\mathbb{E}\left[\sum_{i=1}^{n}\Phi_{i,1}^2 w_{i,j}^2(X_i, X_j')\left(\Psi_{j,1}'\right)^2\Big|X_j', Y_j'\right] \leq T_\psi^2\sup_x\sum_{i=1}^{n}\mathbb{E}\left[w_{i,j}^2(X_i, x)\sigma_{i,\phi}^2(X_i)\right].$$

Combining these two inequalities implies that

$$C \leq T_\phi\sup_x\left(\sum_{j=1}^{n}\mathbb{E}\left[w_{i,j}^2(x, X_j)\sigma_{j,\psi}^2(X_j)\right]\right)^{1/2} + T_\psi\sup_x\left(\sum_{i=1}^{n}\mathbb{E}\left[w_{i,j}^2(X_i, x)\sigma_{i,\phi}^2(X_i)\right]\right)^{1/2}$$

Finally, it is clear from assumption **(A2)** that $D \leq T_\phi T_\psi B_w$. Combining all these with Theorem 3.2 of Giné et al. (2000) and noting

$$T_\phi \leq K_\alpha C_\phi(\log n)^{1/\alpha} \quad \text{and} \quad T_\psi \leq K_\beta C_\psi(\log n)^{1/\beta},$$

we get that there exists a constant $K > 0$ such that for all $p \geq 1$

$$\left\|\mathcal{U}_n^{(1)}\right\|_p \leq K\left[\sqrt{p}\Lambda_{1/2} + p\Lambda_1 + p^{3/2}\left\{\Lambda_{3/2}^{(\alpha)} + \Lambda_{3/2}^{(\beta)}\right\} + p^2\Lambda_2\right]. \tag{12}$$

To bound $\mathcal{U}_n^{(2)}$ and $\mathcal{U}_n^{(3)}$ in (22), we use Hoffmann-Jøgensen's inequality (Proposition 6.8 of Ledoux and Talagrand (1991)). Observe that

$$\mathcal{U}_n^{(2)} := \sum_{i=1}^{n}\varepsilon_i\Phi_{i,2}g_i(X_i;\mathcal{Z}_n'), \quad \text{where} \quad g_i(X_i;\mathcal{Z}_n') := \sum_{j=1, j\neq i}^{n} w_{i,j}(X_i, X_j')\Psi_{j,1}'\varepsilon_j'.$$

With $\mathcal{Z}_n' := \{(\varepsilon_1', Z_1'), \ldots, (\varepsilon_n', Z_n')\}$ and $\mathcal{X}_n := \{X_1, \ldots, X_n\}$, note that

$$\mathbb{P}\left(\max_{1\le I\le n}\left|\sum_{i=1}^{I}\varepsilon_i\Phi_{i,2}g_i(X_i,\mathcal{Z}_n')\right| > 0\big|\mathcal{X}_n,\mathcal{Z}_n'\right) \le \mathbb{P}\left(\max_{1\le i\le n}|\phi(Z_i)| \ge T_\phi\big|\mathcal{X}_n\right) \le 1/8,$$

and so, by Equation (6.8) of Ledoux and Talagrand (1991), we get

$$\begin{aligned}\mathbb{E}\left[\mathcal{U}_n^{(2)}\big|\mathcal{X}_n,\mathcal{Z}_n'\right] &\le 8\mathbb{E}\left[\max_{1\le i\le n}\left|\Phi_{i,2}\left(g_i(X_i;\mathcal{Z}_n')\right)\right|\big|\mathcal{X}_n,\mathcal{Z}_n'\right]\\ &\le 8\mathbb{E}\left[\max_{1\le i\le n}|\phi(Z_i)|\big|\mathcal{X}_n\right]\max_{1\le i\le n}\left|g_i(X_i;\mathcal{Z}_n')\right| = T_\phi\max_{1\le i\le n}\left|g_i(X_i;\mathcal{Z}_n')\right|.\end{aligned}$$

From assumption **(A1)** and Theorem 6.21 of Ledoux and Talagrand (1991), we thus get for $0<\alpha\le 1$,

$$\begin{aligned}\left\|\mathcal{U}_n^{(2)}\right\|_{\psi_\alpha|\mathcal{X}_n,\mathcal{Z}_n'} &\le K_\alpha\mathbb{E}\left[\mathcal{U}_n^{(2)}\big|\mathcal{X}_n,\mathcal{Z}_n'\right] + K_\alpha\left\|\max_{1\le i\le n}\left|\Phi_{i,2}g_i(X_i;\mathcal{Z}_n')\right|\right\|_{\psi_\alpha|\mathcal{X}_n,\mathcal{Z}_n'}\\ &\le K_\alpha\left(T_\phi + \left\|\max_{1\le i\le n}|\phi(Y_i)|\right\|_{\psi_\alpha|\mathcal{X}_n,\mathcal{Z}_n'}\right)\max_{1\le i\le n}\left|g_i(X_i;\mathcal{Z}_n')\right| \\ &\le K_\alpha C_\phi(\log n)^{1/\alpha}\max_{1\le i\le n}\left|g_i(X_i;\mathcal{Z}_n')\right|,\end{aligned}\tag{13}$$

for some constant K_α depending only on α (and can be different in different lines). If $\alpha\ge 1$, then we get

$$\left\|\mathcal{U}_n^{(2)}\right\|_{\psi_{\alpha^*}|\mathcal{X}_n,\mathcal{Z}_n'} \le K_\alpha C_\phi(\log n)^{1/\alpha}\max_{1\le i\le n}\left|g_i(X_i;\mathcal{Z}_n')\right|.$$

See proof of Theorem 3.3 in Kuchibhotla and Chakrabortty (2018) for similar argument. Thus,

$$\mathbb{E}\left[|\mathcal{U}_n^{(2)}|^p\big|\mathcal{X}_n,\mathcal{Z}_n'\right] \le K_\alpha^p C_\phi^p(\log n)^{p/\alpha}p^{p/\alpha^*}\max_{1\le i\le n}\left|g_i(X_i;\mathcal{Z}_n')\right|^p.$$

Thus, for $p\ge 1$,

$$\mathbb{E}\left[|\mathcal{U}_n^{(2)}|^p\right] \le K_\alpha^p C_\phi^p(\log n)^{p/\alpha}p^{p/\alpha^*}\mathbb{E}\left[\max_{1\le i\le n}\left|g_i(X_i;\mathcal{Z}_n')\right|^p\right].\tag{14}$$

To control the right hand side above, recall that

$$g_i(x;\mathcal{Z}_n') = \sum_{j=1,j\neq i}^{n} w_{i,j}(x,X_j')\Psi_{j,1}'\varepsilon_j',$$

is a sum of mean zero independent random variables that are bounded by B_wT_ψ. Also, note that

$$\mathrm{Var}(g_i(x;\mathcal{Z}_n')) = \sum_{j=1,j\neq i}^{n}\mathbb{E}\left[w_{i,j}^2(x,X_j')\psi^2(Z_j')\right] = \sum_{j=1,j\neq i}^{n}\mathbb{E}[w_{i,j}^2(x,X_j')\sigma_{j,\psi}^2(X_j')].$$

Therefore by Bernstein's inequality (Lemma 4 of van de Geer and Lederer (2013)), we get that

$$\mathbb{P}\left(\max_{1\le i\le n}|g_i(X_i;\mathcal{Z}_n')| - \Upsilon_\psi\sqrt{6\log(1+n)} - 3B_wT_\psi\log n \ge \Upsilon_\psi\sqrt{t} + 3B_wT_\psi t\right) \le 2e^{-t},$$

where

$$\Upsilon_\psi^2 := \max_x \sum_{j=1,j\neq i}^n \mathbb{E}[w_{i,j}^2(x,X_j')\sigma_{j,\psi}^2(X_j')].$$

So, by Propositions A.3 and A.4 of Kuchibhotla and Chakrabortty (2018), we get that for $p\ge 1$,

$$\mathbb{E}\left[\max_{1\le i\le n}|g_i(X_i;\mathcal{Z}_n')|^p\right] \le C^p\left[(\log n)^{p/2}\Upsilon_\psi^p + (B_wT_\psi)^p(\log n)^p + p^{p/2}\Upsilon_\psi^p + p^p(B_wT_\psi)^p\right].$$

Hence for $p\ge 1$,

$$\begin{aligned}\mathbb{E}\left[|\mathcal{U}_n^{(2)}|^p\right] &\le K_\alpha^pC_\phi^p(\log n)^{p/\alpha}p^{p/\alpha^*}\left[(\log n)^{p/2}\Upsilon_\psi^p + (B_wT_\psi)^p(\log n)^p\right]\\ &\quad + K_\alpha^pC_\phi^p(\log n)^{p/\alpha}p^{p/\alpha^*}\left[p^{p/2}\Upsilon_\psi^p + p^p(B_wT_\psi)^p\right].\end{aligned} \tag{15}$$

A similar calculation for $\mathcal{U}_n^{(3)}$ shows that for $p\ge 1$,

$$\begin{aligned}\mathbb{E}\left[|\mathcal{U}_n^{(3)}|^p\right] &\le K_\beta^pC_\psi^p(\log n)^{p/\beta}p^{p/\beta^*}\left[(\log n)^{p/2}\Upsilon_\phi^p + (B_wT_\phi)^p(\log n)^p\right]\\ &\quad + K_\beta^pC_\psi^p(\log n)^{p/\beta}p^{p/\beta^*}\left[p^{p/2}\Upsilon_\phi^p + p^p(B_wT_\phi)^p\right],\end{aligned} \tag{16}$$

where

$$\Upsilon_\phi^2 := \max_x \sum_{i=1,i\neq j}^n \mathbb{E}[w_{i,j}^2(X_i,x)\sigma_{i,\phi}^2(X_i)].$$

To control $\mathcal{U}_n^{(4)}$, recall that

$$\mathcal{U}_n^{(4)} = \sum_{i=1}^n \varepsilon_i\Phi_{i,2}\left(\sum_{j=1,j\neq i}^n w_{i,j}(X_i,X_j')\Psi_{j,2}'\varepsilon_j'\right).$$

Following the arguments leading to (13), we have

$$\left\|\mathcal{U}_n^{(4)}\right\|_{\psi_{\alpha^*}|\mathcal{X}_n,\mathcal{Z}_n'} \le K_\alpha C_\phi(\log n)^{1/\alpha}\max_{1\le i\le n}\left|\sum_{j=1,j\neq i}^n w_{i,j}(X_i,X_j')\Psi_j'\varepsilon_j'\right|.$$

Conditioning on $\mathcal{X}_n,\mathcal{X}_n'$, the right hand side satisfies the hypothesis of (6.8) of Ledoux and Talagrand (1991) and so by Theorem 6.21 of Ledoux and Talagrand (1991), we get

$$\begin{aligned}\left\|\max_{1\le i\le n}\left|\sum_{j=1,j\neq i}^n w_{i,j}(X_i,X_j')\Psi_j'\varepsilon_j'\right|\right\|_{\psi_{\beta^*}|\mathcal{X}_n,\mathcal{X}_n'} &\le K_\beta\left\|\max_{1\le i\neq j\le n}\left|w_{i,j}(X_i,X_j')\psi(Z_j')\right|\right\|_{\psi_\beta|\mathcal{X}_n,\mathcal{X}_n'}\\ &\le K_\beta B_w(\log n)^{1/\beta}C_\psi,\end{aligned}$$

for some constant K_β depending only on β. Therefore, for $p \geq 1$

$$\mathbb{E}\left[|\mathcal{U}_n^{(4)}|^p\right] \leq K^p C_\psi^p C_\phi^p p^{p(1/\alpha^*+1/\beta^*)} (\log n)^{p(\alpha^{-1}+\beta^{-1})} B_w^p \leq K^p p^{p(1/\alpha^*+1/\beta^*)} \Lambda_2^p, \qquad (17)$$

for some constant $K > 0$.

Combining bounds (15) and (16), we get that for some constant $K > 0$ and for all $p \geq 1$,

$$\begin{aligned}\left\|\mathcal{U}_n^{(2)} + \mathcal{U}_n^{(3)}\right\|_p &\leq Kp^{1/\alpha^*}(\log n)^{1/2}\Lambda_{3/2}^{(\alpha)} + Kp^{1/\beta^*}(\log n)^{1/2}\Lambda_{3/2}^{(\beta)} \\ &\quad + Kp^{1/2+1/\alpha^*}\Lambda_{3/2}^{(\alpha)} + Kp^{1/2+1/\beta^*}\Lambda_{3/2}^{(\beta)} \\ &\quad + K(\log n)\Lambda_2[p^{1/\alpha^*} + p^{1/\beta^*}] + K\Lambda_2[p^{1+1/\alpha^*} + p^{1+1/\beta^*}].\end{aligned}$$

Combining this inequality with (12) and (17), we get for all $p \geq 1$

$$\begin{aligned}\left\|\sum_{\ell=1}^{4}\mathcal{U}_n^{(\ell)}\right\|_p &\leq K\left[p^{1/2}\Lambda_{1/2} + p\Lambda_1 + p^{3/2}\left\{\Lambda_{3/2}^{(\alpha)} + \Lambda_{3/2}^{(\beta)}\right\} + p^2\Lambda_2\right] \\ &\quad + Kp^{1/\alpha^*}(\log n)^{1/2}\Lambda_{3/2}^{(\alpha)} + Kp^{1/\beta^*}(\log n)^{1/2}\Lambda_{3/2}^{(\beta)} \\ &\quad + Kp^{1/2+1/\alpha^*}\Lambda_{3/2}^{(\alpha)} + Kp^{1/2+1/\beta^*}\Lambda_{3/2}^{(\beta)} \\ &\quad + K(\log n)\Lambda_2[p^{1/\alpha^*} + p^{1/\beta^*}] + K\Lambda_2[p^{1+1/\alpha^*} + p^{1+1/\beta^*}] \\ &\quad + Kp^{(1/\alpha^*+1/\beta^*)}\Lambda_2.\end{aligned}$$

Since $\alpha^* \leq 1$ and $\beta^* \leq 1$, we have

$$\min\{p^{1/2+1/\alpha^*}, p^{1/2+1/\beta^*}\} \geq p^{3/2} \text{ and } \min\{p^{1+1/\alpha^*}, p^{1+1/\beta^*}, p^{1/\alpha^*+1/\beta^*}\} \geq p^2.$$

Using these inequalities, the bound above can be simplified as

$$\begin{aligned}\left\|\sum_{\ell=1}^{4}\mathcal{U}_n^{(\ell)}\right\|_p &\leq Kp^{1/2}\Lambda_{1/2} + Kp\Lambda_1 \\ &\quad + Kp^{1/\alpha^*}\left[(\log n)^{1/2}\Lambda_{3/2}^{(\alpha)} + (\log n)\Lambda_2\right] + Kp^{1/\beta^*}\left[(\log n)^{1/2}\Lambda_{3/2}^{(\beta)} + (\log n)\Lambda_2\right] \\ &\quad + Kp^{1/2+1/\alpha^*}\Lambda_{3/2}^{(\alpha)} + Kp^{1/2+1/\beta^*}\Lambda_{3/2}^{(\beta)} \\ &\quad + Kp^{1/\alpha^*+1/\beta^*}\Lambda_2.\end{aligned}$$

Here the constant $K > 0$ depends only on α, β. This completes the proof based on Lemma 1.

A.2 Auxiliary Lemmas Used in Theorem 1

The two lemmas to follow in this section provide explicit (but not necessarily optimal) constants for Equations (3.1) and (2.6) of Giné et al. (2000). These lemmas can be used in the proof of Theorem 3.2 of Giné et al. (2000) to get explicit constants. In this respect, we note that Theorem 3.4.8 of Giné and Nickl (2016) (which was first proved in Houdré and Reynaud-Bouret (2003)) does not imply Theorem 3.2 of Giné et al. (2000) since the result of Giné et al. (2000) applies for unbounded kernels in U-statistics while the result of Giné and Nickl (2016) applies exclusively for bounded kernel U-statistics.

Lemma 2. *Suppose $Z_1, \ldots, Z_n$ are independent mean zero random variables. Then for $p \geq 1$,*

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} Z_i\right|^p\right] \leq 4^p p^{p/2}\left(\sum_{i=1}^{n}\mathbb{E}\left[Z_i^2\right]\right)^{p/2} + 4^p p^p \mathbb{E}\left[\max_{1\leq i\leq n}|Z_i|^p\right].$$

Proof. By Theorem 7 of Boucheron et al. (2005), we get for $p \geq 2$,

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} Z_i\right|^p\right] \leq 2^{p+1}\left(\frac{2p}{e-\sqrt{e}}\right)^{p/2}\mathbb{E}\left[\left(\sum_{i=1}^{n} Z_i^2\right)^{p/2}\right].$$

By Theorem 8 of Boucheron et al. (2005), we get for $p \geq 2$,

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} Z_i^2\right)^{p/2}\right] \leq 3^{p/2}\left(\sum_{i=1}^{n}\mathbb{E}\left[Z_i^2\right]\right)^{p/2} + \left(\frac{3p\kappa}{2}\right)^{p/2}\mathbb{E}\left[\max_{1\leq i\leq n}|Z_i|^p\right],$$

for $\kappa = 0.5\sqrt{e}/(\sqrt{e}-1)$. Thus for $p \geq 2$,

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} Z_i\right|^p\right] \leq 4^p p^{p/2}\left(\sum_{i=1}^{n}\mathbb{E}\left[Z_i^2\right]\right)^{p/2} + 4^p p^p \mathbb{E}\left[\max_{1\leq i\leq n}|Z_i|^p\right].$$

Since the inequality holds true for $p = 1$ trivially, the result follows. □

Lemma 3. *Suppose $\xi_i, 1 \leq i \leq n$ are independent random variables, then for $p \geq 1$ and $\alpha > 0$,*

$$p^{p\alpha}\sum_{i=1}^{n}\mathbb{E}\left[|\xi_i|^p\right] \leq 4(1.5)^{p\alpha}p^{p\alpha}\mathbb{E}\left[\max_{1\leq i\leq n}|\xi_i|^p\right] + 2(1.5)^{p\alpha}\left(\sum_{i=1}^{n}\mathbb{E}\left[|\xi_i|\right]\right)^p.$$

Proof. Fix $p \geq 1$. Define $\delta_0 \geq 0$ such that

$$\delta_0 := \inf\left\{t > 0 : \sum_{i=1}^{n}\mathbb{P}\left(|\xi_i| > t\right) \leq 1\right\}.$$

By (1.4.4) of de la Peña and Giné (1999), it follows that

$$\frac{1}{2}\max\left\{\delta_0^p, \sum_{i=1}^{n}\mathbb{E}\left[|\xi_i|^p \mathbb{1}_{\{|\xi_i|>\delta_0\}}\right]\right\} \leq \mathbb{E}\left[\max_{1\leq i\leq n}|\xi_i|^p\right]. \tag{18}$$

Observe that

$$
\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|^p\right] &= \sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|^p \mathbb{1}_{\{|\xi_i|>\delta_0\}}\right] + \sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|^p \mathbb{1}_{\{|\xi_i|\le\delta_0\}}\right] \\
&\overset{(a)}{\le} 2\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^p\right] + \sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|^p \mathbb{1}_{\{|\xi_i|\le\delta_0\}}\right] \\
&\le 2\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^p\right] + \delta_0^{p-1} \sum_{i=1}^{n} \mathbb{E}\left[|\xi_i| \mathbb{1}_{\{|\xi_i|\le\delta_0\}}\right] \\
&\overset{(a)}{\le} 2\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^p\right] + 2\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^{p-1}\right] \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i| \mathbb{1}_{\{|\xi_i|\le\delta_0\}}\right]\right) \\
&\le 2\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^p\right] + 2\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^{p-1}\right] \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|\right]\right).
\end{aligned}
$$

Inequality (a) follows from (18). To prove the result now, we consider two cases:

– *Case 1:* If

$$
p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^p\right] \le \left(\sum_{i=1}^{n} \mathbb{E}[|\xi_i|]\right)^p,
$$

then

$$
\begin{aligned}
\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^{p-1}\right] \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|\right]\right) &\le \left(\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^p\right]\right)^{(p-1)/p} \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|\right]\right) \\
&\le \frac{1}{p^{(p-1)\alpha}} \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|\right]\right)^{p-1} \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|\right]\right) \\
&\le \frac{1}{p^{(p-1)\alpha}} \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|\right]\right)^{p}.
\end{aligned}
$$

Therefore (in case 1),

$$
p^{p\alpha} \sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|^p\right] \le 2p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^p\right] + 2p^{\alpha} \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|\right]\right)^p. \tag{19}
$$

– *Case 2:* If

$$
p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n} |\xi_i|^p\right] \ge \left(\sum_{i=1}^{n} \mathbb{E}\left[|\xi_i|\right]\right)^p,
$$

then

$$
\begin{aligned}
&\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^{p-1}\right]\left(\sum_{i=1}^n \mathbb{E}\left[|\xi_i|\right]\right)\\
&\qquad\le\left(\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]\right)^{(p-1)/p}\left(\sum_{i=1}^n \mathbb{E}\left[|\xi_i|\right]\right)\\
&\qquad\le\left(\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]\right)^{(p-1)/p} p^{\alpha}\left(\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]\right)^{1/p}\\
&\qquad\le p^{\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right].
\end{aligned}
$$

Therefore (in case 2),

$$
\begin{aligned}
p^{p\alpha}\sum_{i=1}^n \mathbb{E}\left[|\xi_i|^p\right] &\le 2p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]+2p^{\alpha(p+1)}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]\\
&\le 2p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]+2p^{p\alpha}\left(e^{1/e}\right)^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]\\
&\le (2+(1.5)^{p\alpha})p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]. \qquad (20)
\end{aligned}
$$

Combining inequalities (19) and (20), we get for $p\ge 1$ and $\alpha>0$ that

$$
\begin{aligned}
p^{p\alpha}\sum_{i=1}^n \mathbb{E}\left[|\xi_i|^p\right] &\le (2+(1.5)^{p\alpha})p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]+2p^{\alpha}\left(\sum_{i=1}^n \mathbb{E}\left[|\xi_i|\right]\right)^p\\
&\le 4(1.5)^{p\alpha}p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]+2p^{\alpha}\left(\sum_{i=1}^n \mathbb{E}\left[|\xi_i|\right]\right)^p\\
&\le 4(1.5)^{p\alpha}p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]+2(p^{1/p})^{p\alpha}\left(\sum_{i=1}^n \mathbb{E}\left[|\xi_i|\right]\right)^p\\
&\le 4(1.5)^{p\alpha}p^{p\alpha}\mathbb{E}\left[\max_{1\le i\le n}|\xi_i|^p\right]+2(1.5)^{p\alpha}\left(\sum_{i=1}^n \mathbb{E}\left[|\xi_i|\right]\right)^p.
\end{aligned}
$$

This proves the result. □

Lemma 4. *Under the notation of Theorem 1, the quantity B defined in (11) satisfies*

$$
\begin{aligned}
B\le\sup\Bigg\{&\sum_{1\le i\ne j\le n}\mathbb{E}\left[q_i(X_i)\sigma_{i,\phi}(X_i)w_{i,j}(X_i,X_j)\sigma_{j,\psi}(X_j)p_j(X_j)\right]:\\
&\sum_{j=1}^n \mathbb{E}\left[q_i^2(X_i)\right]\le 1,\ \sum_{i=1}^n \mathbb{E}\left[p_j^2(X_j)\right]\le 1\Bigg\}.
\end{aligned}
$$

Proof. Following the proof of Theorem 3.2 of Giné et al. (2000), the quantity B is the square root of the wimpy variance of

$$S_n := \left(\sum_{i=1}^{n} \mathbb{E}\left[F_i^2(\varepsilon_i, Z_i; \mathcal{Z}_n')\middle|\mathcal{Z}_n'\right]\right)^{1/2},$$

where $\mathcal{Z}_n' := \{(\varepsilon_1', Z_1'), \ldots, (\varepsilon_n', Z_n')\}$ and

$$F_i(\varepsilon_i, Z_i; \mathcal{Z}_n') := \varepsilon_i \Phi_{i,1} \sum_{j=1, j\neq i}^{n} w_{i,j}(X_i, X_j')\Psi_{j,1}'\varepsilon_j'.$$

This implies that

$$S_n \le \left(\sum_{i=1}^{n} \mathbb{E}\left[G_i^2(X_i; \mathcal{Z}_n')\middle|\mathcal{Z}_n'\right]\right)^{1/2},$$

where for $\sigma_{i,\phi}^2(x) := \mathbb{E}\left[\phi^2(Y_i)\middle|X_i = x\right]$,

$$G_i(X_i; \mathcal{Z}_n') := \sigma_{i,\phi}(X_i) \sum_{j=1, j\neq i}^{n} w_{i,j}(X_i, X_j')\Psi_{j,1}'\varepsilon_j'.$$

Note that $\sigma_{i,\phi}(\cdot)$ depends on i since the random variables are allowed to be non-identically distributed. Now observe that

$$S_n = \sup\left\{\sum_{i=1}^{n} \int q_i(x)G_i(x; \mathcal{Z}_n')P_{X_i}(dx) : \sum_{i=1}^{n} \int q_i^2(x)P_{X_i}(dx) \le 1\right\}. \tag{21}$$

To prove this, note that for any $\{q_i(\cdot) : 1 \le i \le n\}$ satisfying the (integral) constraint,

$$\begin{aligned}
&\sum_{i=1}^{n} \int q_i(x)G_i(x; \mathcal{Z}_n')P_{X_i}(dx)\\
&\quad\le \sum_{i=1}^{n} \left(\int q_i^2(x)P_{X_i}(dx)\right)^{1/2} \left(\int G_i^2(x; \mathcal{Z}_n')P_{X_i}(dx)\right)^{1/2}\\
&\quad\le \left(\sum_{i=1}^{n} \int q_i^2(x)P_{X_i}(dx)\right)^{1/2} \left(\sum_{i=1}^{n} \int G_i^2(x; \mathcal{Z}_n')P_{X_i}(dx)\right)^{1/2} \le S_n.
\end{aligned}$$

To prove the reverse inequality, define for $1 \le i \le n$,

$$q_i(x) := G_i(x; \mathcal{Z}_n') \left(\sum_{i=1}^{n} \int G_i^2(x; \mathcal{Z}_n')P_{X_i}(dx)\right)^{1/2}.$$

It is clear that $\{q_i(\cdot) : 1 \le i \le n\}$ satisfy the integral constraint in (21) and

$$\sum_{i=1}^{n} \int q_i(x)G_i(x; \mathcal{Z}_n')P_{X_i}(dx) = S_n.$$

This completes the proof of (21). Rewriting the representation (21), we get

$$S_n = \sup_{\sum_{i=1}^n \int q_i^2(x)P_{X_i}(dx)\le 1} \sum_{j=1}^n \varepsilon'_j \Psi'_{j,1}\left(\sum_{i=1,i\neq j}^n \int q_i(x)\sigma_{i,\phi}(x)w_{i,j}(x,X'_j)P_{X_i}(dx)\right).$$

This representation shows that S_n is indeed the supremum of an empirical process. The wimpy variance of this supremum is given by

$$\begin{aligned}
&\sup_{\{q_i(\cdot)\}} \mathrm{Var}\left(\sum_{j=1}^n \varepsilon'_j \Psi'_{j,1}\left(\sum_{i=1,i\neq j}^n \int q_i(x)\sigma_{i,\phi}(x)w_{i,j}(x,X'_j)P_{X_i}(dx)\right)\right)\\
&\quad\le \sup_{\{q_i(\cdot)\}} \sum_{j=1}^n \mathbb{E}\left[\sigma_{j,\psi}^2(X'_j)\left(\sum_{i=1,i\neq j}^n \int q_i(x)\sigma_{i,\phi}(x)w_{i,j}(x,X'_j)P_{X_i}(dx)\right)^2\right]\\
&\quad= \sup_{\{q_i(\cdot)\}} \sum_{j=1}^n \mathbb{E}\left[\sigma_{j,\psi}^2(X_j)\left(\sum_{i=1,i\neq j}^n \int q_i(x)\sigma_{i,\phi}(x)w_{i,j}(x,X_j)P_{X_i}(dx)\right)^2\right].
\end{aligned}$$

Now a duality argument implies that

$$\begin{aligned}
&\sup_{\{q_i(\cdot)\}} \left(\sum_{j=1}^n \mathbb{E}\left[\sigma_{j,\psi}^2(X_j)\left(\sum_{i=1,i\neq j}^n \int q_i(x)\sigma_{i,\phi}(x)w_{i,j}(x,X_j)P_{X_i}(dx)\right)^2\right]\right)^{1/2}\\
&\quad= \sup\Bigg\{\sum_{1\le i\neq j\le n} \mathbb{E}\left[q_i(X_i)\sigma_{i,\phi}(X_i)w_{i,j}(X_i,X_j)\sigma_{j,\psi}(X_j)p_j(X_j)\right]:\\
&\qquad\qquad \sum_{i=1}^n \mathbb{E}\left[p_j^2(X_j)\right]\le 1, \sum_{j=1}^n \mathbb{E}\left[q_i^2(X_i)\right]\le 1\Bigg\}.
\end{aligned}$$

Thus the result follows. $\square$

B Proofs of Results in Section 3

B.1 Proof of Theorem 2

Similar to $\mathcal{U}_n^{(\ell)}, 1 \le \ell \le 4$ defined in the proof of Theorem 1, we define

$$\begin{aligned}
\mathcal{U}_n^{(1)}(\mathcal{W}) &:= \sup_{w\in\mathcal{W}} \left| \sum_{1\le i\neq j\le n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \Psi_{j,1}' \varepsilon_j' \right|, \\
\mathcal{U}_n^{(2)}(\mathcal{W}) &:= \sup_{w\in\mathcal{W}} \left| \sum_{1\le i\neq j\le n} \varepsilon_i \Phi_{i,2} w_{i,j}(X_i, X_j') \Psi_{j,1}' \varepsilon_j' \right|, \\
\mathcal{U}_n^{(3)}(\mathcal{W}) &:= \sup_{w\in\mathcal{W}} \left| \sum_{1\le i\neq j\le n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \Psi_{j,2}' \varepsilon_j' \right|, \\
\mathcal{U}_n^{(4)}(\mathcal{W}) &:= \sup_{w\in\mathcal{W}} \left| \sum_{1\le i\neq j\le n} \varepsilon_i \Phi_{i,2} w_{i,j}(X_i, X_j') \Psi_{j,2}' \varepsilon_j' \right|.
\end{aligned} \tag{22}$$

As in the proof of Theorem 1, we will control each of the terms separately in the following lemmas. All the lemmas below assume **(A1)** and **(A2′)**.

Lemma 5 (Control of $\mathcal{U}_n^{(4)}(\mathcal{W})$)**.** *There exists a constant $K > 0$ (depending only on α, β) such that for all $p \ge 1$,*

$$\left\| \mathcal{U}_n^{(4)}(\mathcal{W}) \right\|_p \le K \Lambda_2(\mathcal{W}) p^{1/\alpha^* + 1/\beta^*}.$$

Proof. Since $\|w_{i,j}\|_\infty \le B_{\mathcal{W}}$ for all $w \in \mathcal{W}$, it follows that

$$\mathcal{U}_n^{(4)}(\mathcal{W}) \le B_{\mathcal{W}} \sum_{1\le i\neq j\le n} |\Phi_{i,2}\Psi_{j,2}'| \le B_{\mathcal{W}} \left(\sum_{i=1}^n |\Phi_{i,2}| \right) \left(\sum_{j=1}^n |\Psi_{j,2}'| \right).$$

By definition

$$\begin{aligned}
\mathbb{P}\left(\max_{1\le I\le n} \sum_{i=1}^I |\Phi_{i,2}| > 0 \big| \mathcal{X}_n \right) &\le \mathbb{P}\left(\max_{1\le i\le n} |\phi(Z_i)| \ge T_\phi \big| \mathcal{X}_n \right) \le 1/8, \\
\mathbb{P}\left(\max_{1\le I\le n} \sum_{i=1}^I |\Psi_{i,2}'| > 0 \big| \mathcal{X}_n' \right) &\le \mathbb{P}\left(\max_{1\le i\le n} |\psi(Z_i')| \ge T_\psi \big| \mathcal{X}_n' \right) \le 1/8.
\end{aligned}$$

Hence by (6.8) of Ledoux and Talagrand (1991), we get that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^n |\Phi_{i,2}| \big| \mathcal{X}_n \right] &\le C\mathbb{E}\left[\max_{1\le i\le n} |\phi(Z_i)| \big| \mathcal{X}_n \right] \\
\mathbb{E}\left[\sum_{i=1}^n |\Psi_{i,2}'| \big| \mathcal{X}_n' \right] &\le C\mathbb{E}\left[\max_{1\le i\le n} |\psi(Z_i')| \big| \mathcal{X}_n' \right],
\end{aligned}$$

for some constant $C > 0$. Thus by applying Theorem 6.21 of Ledoux and Talagrand (1991) to $\sum\{\Phi_{i,1} - \mathbb{E}[\Phi_{i,1}|\mathcal{X}_n]\}$ and $\sum\{\Psi'_{i,2} - \mathbb{E}[\Psi'_{i,2}|\mathcal{X}'_n]\}$, we get

$$\begin{aligned}
\left\|\sum_{i=1}^n |\Phi_{i,2}|\right\|_{\psi_{\alpha^*}|\mathcal{X}_n} &\le C\left\|\max_{1\le i\le n}|\phi(Z_i)|\right\|_{\psi_\alpha|\mathcal{X}_n} \le CC_\phi(\log n)^{1/\alpha},\\
\left\|\sum_{i=1}^n |\Psi'_{i,2}|\right\|_{\psi_{\beta^*}|\mathcal{X}_n} &\le C\left\|\max_{1\le i\le n}|\psi(Z'_i)|\right\|_{\psi_\beta|\mathcal{X}'_n} \le CC_\psi(\log n)^{1/\beta},
\end{aligned}$$

Therefore, for all $p \ge 1$,

$$\left\|\mathcal{U}_n^{(4)}(\mathcal{W})\right\|_p \le KB_{\mathcal{W}}C_\phi C_\psi(\log n)^{\alpha^{-1}+\beta^{-1}}p^{1/\alpha^*+1/\beta^*} = K\Lambda_2(\mathcal{W})p^{1/\alpha^*+1/\beta^*}.$$

This completes the proof. □

The following lemma controls the moments of $\mathcal{U}_n^{(2)}(\mathcal{W})$ and $\mathcal{U}_n^{(3)}(\mathcal{W})$.

Lemma 6 (Control of $\mathcal{U}_n^{(2)}(\mathcal{W})$ and $\mathcal{U}_n^{(3)}(\mathcal{W})$)**.** *There exists a constant $K > 0$ (depending only on α, β) such that for $p \ge 1$,*

$$\begin{aligned}
\left\|\mathcal{U}_n^{(2)}(\mathcal{W})\right\|_p \le\ & Kp^{1/\alpha^*}\left[E_{n,2}(\mathcal{W}) + (\log n)^{1/2}\Sigma_{n,2}^{1/2}(\mathcal{W}) + (\log n)\Lambda_2(\mathcal{W})\right]\\
&+ Kp^{1/2+1/\alpha^*}\Sigma_{n,2}^{1/2}(\mathcal{W}) + Kp^{1+1/\alpha^*}\Lambda_2(\mathcal{W})\\
\left\|\mathcal{U}_n^{(3)}(\mathcal{W})\right\|_p \le\ & Kp^{1/\beta^*}\left[E_{n,1}(\mathcal{W}) + (\log n)^{1/2}\Sigma_{n,1}^{1/2}(\mathcal{W}) + (\log n)\Lambda_2(\mathcal{W})\right]\\
&+ Kp^{1/2+1/\beta^*}\Sigma_{n,1}^{1/2}(\mathcal{W}) + Kp^{1+1/\beta^*}\Lambda_2(\mathcal{W}).
\end{aligned}$$

Proof. We will only prove the bound for $\mathcal{U}_n^{(2)}(\mathcal{W})$ and the proof for $\mathcal{U}_n^{(3)}(\mathcal{W})$ follows very similar arguments. Recall that

$$\mathcal{U}_n^{(2)}(\mathcal{W}) = \sup_{w\in\mathcal{W}}\left|\sum_{i=1}^n \varepsilon_i\Phi_{i,2}g_i(X_i;\mathcal{Z}'_n,w)\right|, \text{ where } g_i(x;\mathcal{Z}'_n,w) := \sum_{j=1,j\ne i}^n \Psi'_{j,1}\varepsilon'_j w_{i,j}(x,X'_j).$$

Here again (6.8) of Ledoux and Talagrand (1991) applies and we get

$$\left\|\mathcal{U}_n^{(2)}(\mathcal{W})\right\|_{\psi_{\alpha^*}|\mathcal{X}_n,\mathcal{Z}'_n} \le KC_\phi(\log n)^{1/\alpha}\max_{1\le i\le n}\sup_{w\in\mathcal{W}}\left|\sum_{j=1,j\ne i}^n \varepsilon'_j\Psi'_{j,1}w_{i,j}(X_i,X'_j)\right|.$$

By a similar calculation, we get

$$\left\|\mathcal{U}_n^{(3)}(\mathcal{W})\right\|_{\psi_{\beta^*}|\mathcal{X}'_n,\mathcal{Z}_n} \le KC_\psi(\log n)^{1/\beta}\max_{1\le j\le n}\sup_{w\in\mathcal{W}}\left|\sum_{i=1,i\ne j}^n \varepsilon_i\Phi_{i,1}w_{i,j}(X_i,X'_j)\right|.$$

Thus, for $p \geq 1$,

$$\mathbb{E}\left[|\mathcal{U}_n^{(2)}(\mathcal{W})|^p\right] \leq K^p C_\phi^p (\log n)^{p/\alpha} p^{p/\alpha^*} \mathbb{E}\left[\max_{1\leq i\leq n}\sup_{w\in\mathcal{W}}\left|\sum_{j=1,j\neq i}^{n} \varepsilon_j' \Psi_{j,1}' w_{i,j}(X_i, X_j')\right|^p\right],$$
$$\mathbb{E}\left[|\mathcal{U}_n^{(3)}(\mathcal{W})|^p\right] \leq K^p C_\psi^p (\log n)^{p/\beta} p^{p/\beta^*} \mathbb{E}\left[\max_{1\leq j\leq n}\sup_{w\in\mathcal{W}}\left|\sum_{i=1,i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j')\right|^p\right]. \tag{23}$$

The right hand side quantities involve supremum of bounded empirical processes for which Talagrand's inequality applies; see proposition 3.1 of Giné et al. (2000). Observe that for any $x \in \mathfrak{X}$,

$$\max_{1\leq j\leq n}\sup_{w\in\mathcal{W}}|\Psi_{j,1}' w_{i,j}(x, X_j')| \leq C_\psi (\log n)^{1/\beta} B_{\mathcal{W}},$$
$$\max_{1\leq i\leq n}\sup_{w\in\mathcal{W}}|\Phi_{i,1} w_{i,j}(X_i, x)| \leq C_\phi (\log n)^{1/\alpha} B_{\mathcal{W}}.$$

By proposition 3.1 of Giné et al. (2000), we obtain for any $x \in \mathfrak{X}$ and $p \geq 1$,

$$\mathbb{E}\left[\sup_{w\in\mathcal{W}}|g_i(x;\mathcal{Z}_n', w)|^p\right] \leq K^p\left\{\bar{E}_{n,2}^p(\mathcal{W}) + p^{p/2}\bar{\Sigma}_{n,2}^{p/2}(\mathcal{W}) + p^p C_\psi^p (\log n)^{p/\beta} B_{\mathcal{W}}^p\right\},$$

where $\bar{E}_{n,2}(\mathcal{W}) = C_\phi^{-1}E_{n,2}(\mathcal{W})/(\log n)^{1/\alpha}$ and $\bar{\Sigma}_{n,2}^{1/2}(\mathcal{W}) = C_\phi^{-1}\Sigma_{n,2}^{1/2}(\mathcal{W})/(\log n)^{1/\alpha}$. Therefore, by following the argument that lead to (15), we get that

$$\begin{aligned}\mathbb{E}&\left[\max_{1\leq i\leq n}\sup_{w\in\mathcal{W}}|g_i(X_i;\mathcal{Z}_n', w)|^p\right] \\ &\leq K^p\left[\bar{E}_{n,2}^p(\mathcal{W}) + p^{p/2}\bar{\Sigma}_{n,2}^{p/2}(\mathcal{W}) + p^p C_\psi^p(\log n)^{p/\beta}B_{\mathcal{W}}^p\right] \\ &\quad + K^p\left[(\log n)^{p/2}\bar{\Sigma}_{n,2}^{p/2}(\mathcal{W}) + (\log n)^p C_\psi^p (\log n)^{p/\beta} B_{\mathcal{W}}^p\right].\end{aligned} \tag{24}$$

Substituting this in (23), we get

$$\begin{aligned}\mathbb{E}\left[|\mathcal{U}_n^{(2)}(\mathcal{W})|^p\right] &\leq K^p p^{p/\alpha^*}\left[E_{n,2}^p(\mathcal{W}) + p^{p/2}\Sigma_{n,2}^{p/2}(\mathcal{W}) + p^p\Lambda_2^p(\mathcal{W})\right] \\ &\quad + K^p p^{p/\alpha^*}\left[(\log n)^{p/2}\Sigma_{n,2}^{p/2}(\mathcal{W}) + (\log n)^p\Lambda_2^p(\mathcal{W})\right].\end{aligned}$$

By a similar calculation, we get

$$\begin{aligned}\mathbb{E}\left[|\mathcal{U}_n^{(3)}(\mathcal{W})|^p\right] &\leq K^p p^{p/\beta^*}\left[E_{n,1}^p(\mathcal{W}) + p^{p/2}\Sigma_{n,1}^{p/2}(\mathcal{W}) + p^p\Lambda_2^p(\mathcal{W})\right] \\ &\quad + K^p p^{p/\beta^*}\left[(\log n)^{p/2}\Sigma_{n,1}^{p/2}(\mathcal{W}) + (\log n)^p\Lambda_2^p(\mathcal{W})\right].\end{aligned}$$

This completes the proof of the result. □

The following lemma controls the moments of $\mathcal{U}_n^{(1)}(\mathcal{W})$. This is a bounded degenerate U-process and is (usually) the dominating term among the four parts.

Lemma 7 (Control of $\mathcal{U}_n^{(1)}(\mathcal{W})$)**.** *There exists a constant $K > 0$ (depending only on α, β) such that for all $p \geq 1$,*

$$
\begin{aligned}
\left\|\mathcal{U}_n^{(1)}(\mathcal{W})\right\|_p \leq K\mathbb{E}\left[\mathcal{U}_n^{(1)}(\mathcal{W})\right] &+ Kp^{1/2}\left(\mathfrak{W}_{n,1}(\mathcal{W}) + \mathfrak{W}_{n,2}(\mathcal{W})\right) \\
&+ Kp\left(\|(\phi w\psi)_{\mathcal{W}}\|_{2\to 2} + E_{n,1}(\mathcal{W}) + E_{n,2}(\mathcal{W}) + \Sigma_{n,2}^{1/2}(\mathcal{W})\sqrt{\log n} + \Lambda_2(\mathcal{W})\log n\right) \\
&+ Kp^{3/2}\left(\Sigma_{n,1}^{1/2}(\mathcal{W}) + \Sigma_{n,2}^{1/2}(\mathcal{W})\right) + Kp^2\Lambda_2(\mathcal{W}).
\end{aligned}
$$

Proof. Recall that

$$
\mathcal{U}_n^{(1)}(\mathcal{W}) = \sup_{w\in\mathcal{W}}\left|\sum_{i=1}^n \varepsilon_i\Phi_{i,1}g_i(X_i;\mathcal{Z}_n', w)\right|, \text{ where } g_i(X_i;\mathcal{Z}_n', w) := \sum_{j=1, j\neq i}^n \varepsilon_j'\Psi_{j,1}'w_{i,j}(X_i, X_j').
$$

Observe that conditional on $\mathcal{Z}_n'$, $\mathcal{U}_n^{(1)}(\mathcal{W})$ is a bounded empirical process and so Talagrand's inequality applies. Thus by Proposition 3.1 of Giné et al. (2000), we get for $p \geq 1$

$$
\begin{aligned}
\mathbb{E}\left[|\mathcal{U}_n^{(1)}(\mathcal{W})|^p\middle|\mathcal{Z}_n'\right] \leq K^p\left(\mathbb{E}\left[|\mathcal{U}_n^{(1)}(\mathcal{W})|\middle|\mathcal{Z}_n'\right]\right)^p& \\
+ K^pp^{p/2}\sup_{w\in\mathcal{W}}\left(\sum_{i=1}^n\mathbb{E}\left[\Phi_{i,1}^2g_i^2(X_i;\mathcal{Z}_n', w)\middle|\mathcal{Z}_n'\right]\right)^{p/2}& \\
+ K^pp^p\mathbb{E}\left[\max_{1\leq i\leq n}|\Phi_{i,1}|^p\sup_{w\in\mathcal{W}}\left|g_i(X_i;\mathcal{Z}_n', w)\right|^p\middle|\mathcal{Z}_n'\right]&.
\end{aligned}
$$

Therefore, for $p \geq 1$,

$$
\begin{aligned}
\mathbb{E}\left[|\mathcal{U}_n^{(1)}(\mathcal{W})|^p\right] &\leq K^p\mathbb{E}\left(\mathbb{E}\left[|\mathcal{U}_n^{(1)}(\mathcal{W})|\middle|\mathcal{Z}_n'\right]\right)^p \\
&\quad + K^pp^{p/2}\mathbb{E}\left[\sup_{w\in\mathcal{W}}\left(\sum_{i=1}^n\mathbb{E}\left[\sigma_{i,\phi}^2(X_i)g_i^2(X_i;\mathcal{Z}_n', w)\middle|\mathcal{Z}_n'\right]\right)^{p/2}\right] \\
&\quad + K^pp^pC_\phi^p(\log n)^{p/\alpha}\mathbb{E}\left[\max_{1\leq i\leq n}\sup_{w\in\mathcal{W}}\left|g_i(X_i;\mathcal{Z}_n', w)\right|^p\right] \\
&=: K^p\left[\mathbf{I} + \mathbf{II} + \mathbf{III}\right].
\end{aligned}
$$

Controlling **III** : Using (24) from Lemma 6, we get

$$
\begin{aligned}
\mathbf{III} \leq K^pp^p&\left[E_{n,2}^p(\mathcal{W}) + p^{p/2}\Sigma_{n,2}^{p/2}(\mathcal{W}) + p^p\Lambda_2^p(\mathcal{W})\right] \\
&+ K^pp^p\left[(\log n)^{p/2}\Sigma_{n,2}^{p/2}(\mathcal{W}) + (\log n)^p\Lambda_2^p(\mathcal{W})\right].
\end{aligned}
$$

Controlling **II** : To control **II**, we use a technique similar to the one used in Lemma 4. For this note by (21) that for any $w(\cdot,\cdot)$

$$\left(\sum_{i=1}^{n}\int \sigma_{i,\phi}^{2}(x)g_i^2(x;\mathcal{Z}_n',w)P_{X_i}(dx)\right)^{1/2}$$
$$= \sup\left\{\sum_{i=1}^{n}\int q_i(x)\sigma_{i,\phi}(x)g_i(x;\mathcal{Z}_n',w)P_{X_i}(dx) : \sum_{i=1}^{n}\int q_i^2(x)P_{X_i}(dx)\le 1\right\}.$$

Therefore,

$$\mathbf{II} = p^{p/2}\mathbb{E}\left[\sup_{w\in\mathcal{W}}\sup_{\sum_{i=1}^{n}\int q_i^2(x)P_{X_i}(dx)\le 1}\left|\sum_{i=1}^{n}\int q_i(x)\sigma_{i,\phi}(x)g_i(x;\mathcal{Z}_n',w)P_{X_i}(dx)\right|^{p}\right].$$

Now observe that

$$\sum_{i=1}^{n}\int q_i(x)\sigma_{i,\phi}(x)g_i(x;\mathcal{Z}_n',w)P_{X_i}(dx) = \sum_{j=1}^{n}\varepsilon_j'\Psi_{j,1}'\ell_j(X_j';\{q_i\},w),$$

where $\{q_i\}$ represents the sequence $(q_1,\ldots,q_n)$ satisfying $\sum_{i=1}^{n}\int q_i^2(x)P_{X_i}(dx)\le 1\}$ and

$$\ell_j(X_j';\{q_i\},w) := \sum_{i=1,i\neq j}^{n}\int q_i(x)\sigma_{i,\phi}(x)w_{i,j}(x,X_j')P_{X_i}(dx).$$

Thus

$$\mathbf{II} = p^{p/2}\mathbb{E}\left[\sup_{w\in\mathcal{W}}\sup_{\{q_i\}}\left|\sum_{j=1}^{n}\varepsilon_j'\Psi_{j,1}'\ell_j(X_j';\{q_i\},w)\right|^{p}\right].$$

The right hand side is a bounded empirical process and by proposition 3.1 of Giné et al. (2000), we get

$$\begin{aligned}
&\mathbb{E}\left[\sup_{w\in\mathcal{W}}\sup_{\{q_i\}}\left|\sum_{j=1}^{n}\varepsilon_j'\Psi_{j,1}'\ell_j(X_j';\{q_i\},w)\right|^{p}\right]\\
&\quad\le K^p\left(\mathbb{E}\left[\sup_{\{q_i\}}\sup_{w\in\mathcal{W}}\left|\sum_{j=1}^{n}\varepsilon_j'\Psi_{j,1}'\ell_j(X_j';\{q_i\},w)\right|\right]\right)^{p}\\
&\qquad + K^p p^{p/2}\sup_{\{q_i\}}\sup_{w\in\mathcal{W}}\left(\mathrm{Var}\left(\sum_{j=1}^{n}\varepsilon_j'\Psi_{j,1}'\ell_j(X_j';\{q_i\},w)\right)\right)^{p/2}\\
&\qquad + K^p p^{p}\mathbb{E}\left[\sup_{\{q_i\}}\sup_{w\in\mathcal{W}}\max_{1\le j\le n}|\Psi_{j,1}'|^{p}|\ell_j(X_j';\{q_i\},w)|^{p}\right].
\end{aligned}\tag{25}$$

We will now control each of the three terms appearing in (25). Using the fact $|\Psi'_{j,1}| \le KC_\phi(\log n)^{1/\beta}$, we get

$$\mathbb{E}\left[\sup_{\{q_i\}}\sup_{w\in\mathcal{W}}\max_{1\le j\le n}|\Psi'_{j,1}|^p|\ell_j(X'_j;\{q_i\},w)|^p\right] \le C_\psi^p(\log n)^{p/\beta}\sup_{w\in\mathcal{W}}\sup_{x'\in\mathfrak{X}}\sup_{\{q_i\}}|\ell_j(x';\{q_i\},w)|^p.$$

By following the duality argument (21), we get

$$\sup_{\{q_i\}}|\ell_j(x';\{q_i\},w)| \le \left(\sum_{i=1,i\ne j}^{n}\mathbb{E}\left[\sigma_{i,\phi}^2(X_i)w^2(X_i,x')\right]\right)^{1/2},$$

and so,

$$\begin{aligned}&\mathbb{E}\left[\sup_{\{q_i\}}\sup_{w\in\mathcal{W}}\max_{1\le j\le n}|\Psi'_{j,1}|^p|\ell_j(X'_j;\{q_i\},w)|^p\right]\\ &\quad\le K^pC_\psi^p(\log n)^{p/\beta}\sup_{w\in\mathcal{W},x\in\mathfrak{X}}\left(\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{i,\phi}^2(X_i)w^2(X_i,x')\right]\right)^{p/2} = K^p\Sigma_{n,1}^{p/2}(\mathcal{W}).\end{aligned} \tag{26}$$

Also, note that

$$\operatorname{Var}\left(\sum_{j=1}^{n}\varepsilon'_j\Psi'_{j,1}\ell_j(X'_j;\{q_i\},w)\right) = \sum_{j=1}^{n}\mathbb{E}\left[\sigma_{j,\psi}^2(X'_j)\ell_j^2(X'_j;\{q_i\},w)\right].$$

Hence, again following the duality argument (21), we get

$$\begin{aligned}&\sup_{\{q_i\}}\sup_{w\in\mathcal{W}}\left(\operatorname{Var}\left(\sum_{j=1}^{n}\varepsilon'_j\Psi'_{j,1}\ell_j(X'_j;\{q_i\},w)\right)\right)^{p/2}\\ &\quad\le\sup_{\{q_i\}}\sup_{w\in\mathcal{W}}\sup_{\{p_j\}}\left(\sum_{1\le i\ne j\le n}\mathbb{E}\left[q_i(X_i)\sigma_{i,\phi}(X_i)w_{i,j}(X_i,X'_j)\sigma_{j,\psi}(X'_j)p_j(X'_j)\right]\right)^{p}.\end{aligned} \tag{27}$$

Here $\{p_j\}$ represents a sequence $(p_1,\ldots,p_n)$ satisfying $\sum_{j=1}^n\int p_j^2(x)P_{X_j}(dx)\le 1$.

Substituting (27) and (26) in (25), we get

$$\begin{aligned}\mathbf{II} \le K^pp^{p/2}&\left(\mathbb{E}\left[\sup_{\{q_i\}}\sup_{w\in\mathcal{W}}\left|\sum_{j=1}^{n}\varepsilon'_j\Psi'_{j,1}\ell_j(X'_j;\{q_i\},w)\right|\right]\right)^p\\ &+K^pp^p\left\|(\phi w\psi)_{\mathcal{W}}\right\|_{2\to 2}^p + K^pp^{3p/2}\Sigma_{n,1}^{p/2}(\mathcal{W}).\end{aligned}$$

Controlling **I** : We use Lemma 8 (a restatement of Lemma 2 of Adamczak (2006)) to control **I**. In the notation of Lemma 8, take

$$W_j = (\varepsilon'_j, Z'_j),\ T = (Z_1,\ldots,Z_n,\varepsilon_1,\ldots,\varepsilon_n),$$

and for $w \in \mathcal{W}$,

$$f_j^w(W_j, T) = \sum_{i=1, i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \Psi_{j,1}' \varepsilon_j'.$$

This implies

$$S = \mathbb{E}_T \left[\sup_{w\in\mathcal{W}} \left| \sum_{j=1}^{n} \sum_{i=1, i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \Psi_{j,1}' \varepsilon_j' \right| \right] = \mathbb{E}\left[\mathcal{U}_n^{(1)}(\mathcal{W}) \middle| \mathcal{Z}_n' \right].$$

Observe that $\mathbb{E}[S] = \mathbb{E}[\mathcal{U}_n^{(1)}(\mathcal{W})]$. Thus we get for $p \geq 1$

$$\begin{aligned} \mathbb{E}[S^p] &\leq K^p (\mathbb{E}[S])^p + K^p p^{p/2} \Upsilon^p \\ &\quad + K^p p^p \mathbb{E}\left[\max_{1\leq j\leq n} \left(\mathbb{E}\left[\sup_{w\in\mathcal{W}} \left| \sum_{i=1, i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \Psi_{j,1}' \varepsilon_j' \right| \middle| \mathcal{Z}_n' \right] \right)^p \right], \end{aligned} \tag{28}$$

where

$$\Upsilon := \sup_{q\in\mathcal{Q}} \left(\sum_{j=1}^{n} \mathbb{E}\left[\left(\sum_{w\in\mathcal{W}} \mathbb{E}_T\left[f_j^w(W_j, T) q_j(T) \right] \right)^2 \right] \right)^{1/2},$$

with $\mathcal{Q}$ defined in Lemma 8. We now simplify the last two terms on the right hand side of (28). First observe that for the third term

$$\begin{aligned} &\mathbb{E}\left[\sup_{w\in\mathcal{W}} \left| \sum_{i=1, i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \Psi_{j,1}' \varepsilon_j' \right| \middle| \mathcal{Z}_n' \right] \\ &\qquad \leq K C_\psi (\log n)^{1/\beta} \sup_{x\in\mathfrak{X}} \mathbb{E}\left[\sup_{w\in\mathcal{W}} \left| \sum_{i=1, i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, x) \right| \right] = K E_{n,1}(\mathcal{W}). \end{aligned}$$

To control Υ, observe that

$$\sum_{w\in\mathcal{W}} \mathbb{E}_T\left[f_j^w(W_j, T) q_j(T) \right] = \varepsilon_j' \Psi_{j,1}' \sum_{w\in\mathcal{W}} \mathbb{E}_T\left[q_j(T) \sum_{i=1, i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \right].$$

So, using the definition of $\sigma_{j,\psi}^2(\cdot)$, we get

$$\begin{aligned} \Upsilon &= \sup_{q\in\mathcal{Q}} \left(\sum_{j=1}^{n} \mathbb{E}\left[\sigma_{j,\psi}^2(X_j') \left(\sum_{w\in\mathcal{W}} \mathbb{E}_T\left[q_j(T) \sum_{i=1, i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \right] \right)^2 \right] \right)^{1/2} \\ &\overset{(a)}{=} \sup_{\{p_j\}, q\in\mathcal{Q}} \sum_{j=1}^{n} \mathbb{E}\left[p_j(X_j') \sigma_{j,\psi}(X_j') \sum_{w\in\mathcal{W}} \mathbb{E}_T\left[q_j(T) \sum_{i=1, i\neq j}^{n} \varepsilon_i \Phi_{i,1} w_{i,j}(X_i, X_j') \right] \right] \\ &\overset{(b)}{=} \sup_{\{p_j\}} \mathbb{E}\left[\sup_{w\in\mathcal{W}} \sum_{i=1}^{n} \varepsilon_i \Phi_{i,1} \left(\sum_{j=1, j\neq i}^{n} \int p_j(x) \sigma_{j,\psi}(x) w_{i,j}(X_i, x) P_{X_j}(dx) \right) \right]. \end{aligned}$$

Equality (a) above follows from the duality argument (21) while equality (b) follows from the argument given in Lemma 8. $\square$

B.2 Auxiliary Lemmas Used in Theorem 2

The following lemma is a rewording of Lemma 2 of Adamczak (2006). For this result, define the class of functions

$$\mathcal{Q} := \left\{ q(\cdot) = (q_1(\cdot), q_2(\cdot), \dots) : \sum_{k=1}^{\infty} |q_k(T)| = 1 \quad \text{for all} \quad T \right\}.$$

The domain of functions in $\mathcal{Q}$ is left out on purpose.

Lemma 8. *Suppose $\mathcal{F} := \{(f_1^k, \dots, f_n^k) : k \geq 1\}$ represents a countable class of vector functions. Define for independent random variables $T, W_1, \dots, W_n$,*

$$S := \mathbb{E}_T \left[\sup_{k \geq 1} \left| \sum_{j=1}^{n} f_j^k(W_j, T) \right| \right],$$

where $\mathbb{E}_T[\cdot]$ represents the expectation only with respect to T. (So, S is a random variable that depends on $W_1, \dots, W_n$). If $\mathbb{E}_W[f_j^k(W_j, T)] = 0$ for a.e T, then there exists a constant $K > 0$ such that for all $p \geq 1$,

$$\begin{aligned}\mathbb{E}\left[S^p\right] \leq K^p (\mathbb{E}[S])^p + K^p p^{p/2} \sup_{q \in \mathcal{Q}} \left(\sum_{j=1}^{n} \mathbb{E}\left[\left(\sum_{k=1}^{\infty} \mathbb{E}_T[f_j^k(W_j, T) q_j(T)] \right)^2 \right] \right)^{p/2} \\ + K^p p^p \mathbb{E}\left[\max_{1 \leq j \leq n} \left(\mathbb{E}_T \left[\sup_{k \geq 1} |f_j^k(W_j, T)| \right] \right)^p \right].\end{aligned}$$

Proof. Following the proof of Lemma 2 of Adamczak (2006), we get

$$S = \sup_{q \in \mathcal{Q}} \left| \sum_{k=1}^{\infty} \mathbb{E}_T \left[q_k(Y) \sum_{j=1}^{n} f_j^k(W_j, T) \right] \right|.$$

To see this, define $\hat{q}(\cdot) = (\hat{q}_1(\cdot), \dots) \in \mathcal{Q}$ such that

$$\hat{q}_{\hat{k}}(t) = \text{sign}\left(\sum_{j=1}^{n} f_j^{\hat{k}}(W_j, T) \right), \quad \text{and} \quad \hat{q}_k(t) = 0, \quad \text{for } k \neq \hat{k}.$$

Here $\hat{k}$ satisfying

$$\left| \sum_{j=1}^{n} f_j^{\hat{k}}(W_j, T) \right| = \sup_{k \geq 1} \left| \sum_{j=1}^{n} f_j^k(W_j, T) \right|.$$

Therefore,

$$S = \sup_{q\in\mathcal{Q}} \left| \sum_{j=1}^{n} \left(\sum_{k=1}^{\infty} \mathbb{E}_T \left[q_k(T) f_j^k(W_j, T) \right] \right) \right| =: \sup_{q\in\mathcal{Q}} \left| \sum_{j=1}^{n} g_{q,j}(W_j) \right|.$$

The right hand side above is the supremum of a mean zero empirical process and so by proposition 3.1 of Giné et al. (2000), we get

$$\mathbb{E}\left[S^p\right] \le K^p (\mathbb{E}[S])^p + K^p p^{p/2} \sup_{q\in\mathcal{Q}} \left(\sum_{j=1}^{n} \mathbb{E}\left[g_{q,j}^2(W_j) \right] \right)^{p/2} + K^p p^p \mathbb{E}\left[\max_{1\le j\le n} \sup_{q\in\mathcal{Q}} |g_{q,j}(W_j)|^p \right].$$

From the definition of $\mathcal{Q}$, we get

$$\sup_{q\in\mathcal{Q}} |g_{q,j}(W_j)| = \sup_{q\in\mathcal{Q}} \left| \sum_{k=1}^{\infty} \mathbb{E}_T \left[q_k(T) f_j^k(W_j, T) \right] \right| = \mathbb{E}_T \left[\sup_{k\ge 1} |f_j^k(W_j, T)| \right].$$

Thus,

$$\mathbb{E}\left[\max_{1\le j\le n} \sup_{q\in\mathcal{Q}} |g_{q,j}(W_j)|^p \right] = \mathbb{E}\left[\max_{1\le j\le n} \left(\mathbb{E}_T \left[\sup_{k\ge 1} |f_j^k(W_j, T)| \right] \right)^p \right].$$

So, the result follows. □

C Proof of the Maximal Inequality

The following moment bound of Rademacher chaos is used in the proof. See corollary 3.2.6 of de la Peña and Giné (1999) and inequalities leading to (4.1.20) on page 167 of de la Peña and Giné (1999).

Lemma 9. *Let Z be a homogeneous Rademacher chaos of degree 2, that is,*

$$Z := \sum_{1\le i\ne j\le n} \epsilon_i \epsilon_j a_{i,j},$$

for some constants $a_{i,j}, 1 \le i \ne j \le n$. Then $\|Z\|_{\psi_1} \le 4e s_n$, where

$$s_n^2 := \sum_{1\le i\ne j\le n} a_{i,j}^2.$$

Proof of Theorem 3. As before, let $\mathcal{X}_n := \{X_1, X_2, \ldots, X_n\}$. Also, let

$$Z_\epsilon(f) := \left| \frac{1}{\sqrt{n(n-1)}} \sum_{1\le i\ne j\le n} \epsilon_i \epsilon_j f_{i,j}(X_i, X_j) \right|.$$

By Lemma 9, we get conditional on $\mathcal{X}_n$,

$$\|Z_\epsilon(f)\|_{\psi_1|\mathcal{Z}_n} \leq 4e \left(\frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} f_{i,j}^2(X_i, X_j)\right)^{1/2} \leq 4e \, \|f\|_{2,P_n},$$

where

$$\|f\|_{2,P_n} := \left(\frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} f_{i,j}^2(X_i, X_j)\right)^{1/2},$$

and define the discrete probability measure P_n with support $\{X_1, \dots, X_n\}$ as

$$P_n(\{X_i\}) := \frac{1}{n} \quad \text{for} \quad 1 \leq i \leq n.$$

Now, following the proof of Theorem 5.1.4 of de la Peña and Giné (1999),

$$\left\|\max_{f\in\mathcal{F}} Z_\epsilon(f)\right\|_{\psi_1|\mathcal{X}_n} \leq C \int_0^{\Delta_n} \log N\left(\varepsilon, \mathcal{F}, \|\cdot\|_{2,P_n}\right) d\varepsilon,$$

where

$$\Delta_n := \sup_{f\in\mathcal{F}} \|f\|_{2,P_n}.$$

Therefore,

$$\left\|\max_{f\in\mathcal{F}} Z_\epsilon(f)\right\|_{\psi_1|\mathcal{X}_n} \leq C \, \|F\|_{2,P_n} \, J_2\left(\frac{\Delta_n}{\|F\|_{2,P_n}}, \mathcal{F}, \|\cdot\|_2\right).$$

This implies that

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} Z_\epsilon(f)\right] \leq C\mathbb{E}\left[\|F\|_{2,P_n} \, J_2\left(\frac{\Delta_n}{\|F\|_{2,P_n}}, \mathcal{F}, \|\cdot\|_2\right)\right]. \tag{29}$$

Using concavity of $(x, y) \mapsto \sqrt{y} J_2(\sqrt{x/y}, \mathcal{F}, \|\cdot\|_2)$ as in the proof of Theorem 2.1 of van der Vaart and Wellner (2011), it follows that

$$\mathbb{E}\left[\|F\|_{2,P_n} \, J_2\left(\frac{\Delta_n}{\|F\|_{2,P_n}}, \mathcal{F}, \|\cdot\|_2\right)\right] \leq \|F\|_{2,P} \, J_2\left(\frac{\sqrt{\mathbb{E}\left[\Delta_n^2\right]}}{\|F\|_{2,P}}, \mathcal{F}, \|\cdot\|_2\right), \tag{30}$$

where

$$\|F\|_{2,P}^2 := \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \mathbb{E}\left[F_{i,j}^2(X_i, X_j)\right].$$

At this point the proof of Theorem 5.1 of Chen and Kato (2017) uses Hoeffding averaging to bound $\mathbb{E}\left[\Delta_n^2\right]$ which proves the result for iid random variables X_i. To allow for non-identically distributed random variables $X_i, 1 \leq i \leq n$, we bound $\mathbb{E}[\Delta_n^2]$ in terms of J_2

on the right hand side of (30). This is similar to the proof of Theorem 2.1 of van der Vaart and Wellner (2011). To bound $\mathbb{E}\left[\Delta_n^2\right]$, define for $f \in \mathcal{F}$,

$$
\begin{aligned}
W_n^{(1)}(f) := \frac{1}{n(n-1)} \Bigg| \sum_{1\le i\neq j\le n} & \left\{f_{i,j}^2(X_i, X_j) - \mathbb{E}\left[f_{i,j}^2(X_i, X_j)|X_i\right]\right\} \\
& - \left\{\mathbb{E}\left[f_{i,j}^2(X_i, X_j)|X_j\right] + \mathbb{E}\left[f_{i,j}^2(X_i, X_j)\right]\right\}\Bigg|, \\
W_n^{(2)}(f) := \frac{1}{n(n-1)} \Bigg| \sum_{1\le i\neq j\le n} & \left\{\mathbb{E}\left[f_{i,j}^2(X_i, X_j)|X_i\right] - \mathbb{E}\left[f_{i,j}^2(X_i, X_j)\right]\right\}\Bigg|, \\
W_n^{(3)}(f) := \frac{1}{n(n-1)} \Bigg| \sum_{1\le i\neq j\le n} & \left\{\mathbb{E}\left[f_{i,j}^2(X_i, X_j)|X_j\right] - \mathbb{E}\left[f_{i,j}^2(X_i, X_j)\right]\right\}\Bigg|.
\end{aligned}
$$

Using these definitions, we get

$$
\Delta_n^2 \le \sup_{f\in\mathcal{F}} W_n^{(1)}(f) + \sup_{f\in\mathcal{F}} W_n^{(2)}(f) + \sup_{f\in\mathcal{F}} W_n^{(3)}(f) + \Sigma_n^2(\mathcal{F}), \tag{31}
$$

where

$$
\Sigma_n^2(\mathcal{F}) := \sup_{f\in\mathcal{F}} \frac{1}{n(n-1)} \sum_{1\le i\neq j\le n} \mathbb{E}\left[f_{i,j}^2(X_i, X_j)\right].
$$

By decoupling and symmetrization, we obtain

$$
\mathbb{E}\left[\sup_{f\in\mathcal{F}} W_n^{(1)}(f)\right] \le C\mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n(n-1)}\left|\sum_{1\le i\neq j\le n} \epsilon_i\epsilon_j f_{i,j}^2(X_i, X_j)\right|\right].
$$

Set for $f \in \mathcal{F}$,

$$
R_\epsilon(f) := \frac{1}{\sqrt{n(n-1)}} \sum_{1\le i\neq j\le n} \epsilon_i\epsilon_j f_{i,j}^2(X_i, X_j).
$$

Again by Lemma 9 and using $|f_{i,j}(x, x') + g_{i,j}(x, x')| \le 2R$ for all $f, g \in \mathcal{F}$ and $x, x' \in \mathcal{X}$, we get

$$
\begin{aligned}
\left\|R_\epsilon(f) - R_\epsilon(g)\right\|_{\psi_1 \left| \mathcal{X}_n\right.} &\le 8eR\left(\frac{1}{n(n-1)} \sum_{1\le i\neq j\le n} \left(f_{i,j}(X_i, X_j) - g_{i,j}(X_i, X_j)\right)^2\right)^{1/2} \\
&\le 8eR\left\|f - g\right\|_{2,P_n}.
\end{aligned}
$$

Hence by following the first part of the proof, we get

$$
\mathbb{E}\left[\sup_{f\in\mathcal{F}} W_n^{(1)}(f)\right] \le C\frac{R\left\|F\right\|_{2,P}}{n} J_2\left(\frac{\sqrt{\mathbb{E}\left[\Delta_n^2\right]}}{\left\|F\right\|_{2,P}}, \mathcal{F}, \left\|\cdot\right\|_2\right).
$$

Substituting this in (31) after taking expectations,

$$\frac{\|\Delta_n\|_2^2}{\|F\|_{2,P}^2} \le CB_n^2 J_2\left(\frac{\|\Delta_n\|_2}{\|F\|_{2,P}}, \mathcal{F}, \|\cdot\|_2\right) + A_n^2,$$

where

$$B_n^2 := \frac{R}{n\,\|F\|_{2,P}} \quad \text{and} \quad A_n^2 := \frac{\mathbb{E}\left[\sup_{f\in\mathcal{F}} W_n^{(2)}(f)\right] + \mathbb{E}\left[\sup_{f\in\mathcal{F}} W_n^{(3)}(f)\right] + \Sigma_n^2(\mathcal{F})}{\|F\|_{2,P}^2}.$$

It follows that

$$\frac{\|\Delta_n\|_2^2}{\|F\|_{2,P}^2} \le Cb_n^2 J_2\left(\frac{\|\Delta_n\|_2}{\|F\|_{2,P}}, \mathcal{F}, \|\cdot\|_2\right) + a^2,$$

for any $a \ge A_n$ and $b \ge B_n$. Therefore, by Lemma 2.1 of van der Vaart and Wellner (2011), it follows that for any $a \ge A_n$ and $b \ge B_n$,

$$J_2\left(\frac{\|\Delta_n\|_2}{\|F\|_{2,P}}, \mathcal{F}, \|\cdot\|_2\right) \le CJ_2(a, \mathcal{F}, \|\cdot\|_2)\left[1 + \frac{J_2(a, \mathcal{F}, \|\cdot\|_2)b^2}{a^2}\right].$$

Substituting this in (30) and (29), we get

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} Z_\epsilon(f)\right] \le C\,\|F\|_{2,P}\, J_2(a)\left[1 + \frac{J_2(a^2, \mathcal{F}, \|\cdot\|_2)b^2}{a^2}\right],$$

for any $a \ge A_n$ and $b \ge B_n$. The result is proved. □

Bibliography

Adamczak, R. (2006). Moment inequalities for U-statistics. *Ann. Probab.*, 34(6):2288–2314.

Arcones, M. A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Probab.*, 21(3):1494–1542.

Boucheron, S., Bousquet, O., Lugosi, G., and Massart, P. (2005). Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities*. Oxford University Press, Oxford. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

Chen, X. and Kato, K. (2017). Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. *ArXiv e-prints:1708.02705*.

Clémençon, S., Lugosi, G., and Vayatis, N. (2008). Ranking and empirical minimization of U-statistics. *Ann. Statist.*, 36(2):844–874.

de la Peña, V. H. (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.*, 20(4):1877–1892.

de la Peña, V. H. and Giné, E. (1999). *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

Dirksen, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.*, 20:no. 53, 29.

Giné, E., Latała, R., and Zinn, J. (2000). 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.

Giné, E. and Nickl, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli*, 14(1):47–61.

Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.*, 6(2):109–115.

Houdré, C. and Reynaud-Bouret, P. (2003). 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.

Kolesko, K. and Latała, R. (2015). Moment estimates for chaoses generated by symmetric random variables with logarithmically convex tails. *Statist. Probab. Lett.*, 107:210–214.

Kuchibhotla, A. K. and Chakrabortty, A. (2018). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *arXiv preprint arXiv:1804.02605*.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin. Isoperimetry and processes.

Major, P. (2005). Tail behaviour of multiple random integrals and U-statistics. *Probab. Surv.*, 2:448–505.

Major, P. (2013). *On the estimation of multiple random integrals and U-statistics*, volume 2079 of *Lecture Notes in Mathematics*. Springer, Heidelberg.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.*, 15(2):780–799.

Nolan, D., Pollard, D., et al. (1988). Functional limit theorems for u-processes. *The Annals of Probability*, 16(3):1291–1298.

Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18:no. 82, 9.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Spokoiny, V. and Zhilova, M. (2013). Sharp deviation bounds for quadratic forms. *Math. Methods Statist.*, 22(2):100–113.

Talagrand, M. (2014). *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics].* Springer, Heidelberg. Modern methods and classical problems.

van de Geer, S. and Lederer, J. (2013). The Bernstein-Orlicz norm and deviation inequalities. *Probab. Theory Related Fields*, 157(1-2):225–250.

van der Vaart, A. and Wellner, J. A. (2011). A local maximal inequality under uniform entropy. *Electron. J. Stat.*, 5:192–203.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.