Exponential inequalities for dependent V-statistics via random Fourier features

Yandi Shen*    Fang Han*    Daniela Witten*

Abstract

We establish exponential inequalities for a class of V-statistics under strong mixing conditions. Our theory is developed via a novel kernel expansion based on random Fourier features and the use of a probabilistic method. This type of expansion is new and useful for handling many notorious classes of kernels.

Keywords: dependent V-statistics; strong mixing condition; kernel expansion; random Fourier features.

AMS MSC 2010: 60F10.

Submitted to EJP on March 21, 2019, final version accepted on January 5, 2020.

1 Introduction

Consider the following V-statistic of order \( m \) generated by the symmetric kernel \( f \),

\[
V_n(f) := \sum_{i_1, \ldots, i_m = 1}^{n} f(X_{i_1}, \ldots, X_{i_m}),
\]

where \( \{X_i\}_{i=1}^{n} \) is a stationary sequence with marginal measure \( P \) on the \( d \)-dimensional real space \( \mathbb{R}^d \). The purpose of this paper is to establish exponential-type tail bounds for (1.1) when \( \{X_i\}_{i=1}^{n} \) are weakly dependent.

In (1.1), if the summation is taken over \( m \)-tuples \((i_1, \ldots, i_m)\) of distinct indices, the resulting is a U-statistic. In many applications, the techniques of analyzing U- and V-statistics are the same. Non-asymptotic tail bounds and limiting theorems of V- and U-statistics in the i.i.d. case have also been extensively studied [17, 2, 15, 1].

The analysis of V- and U-statistics when the observed data are no longer independent has attracted increasing attention in statistics and probability, with most of the efforts put on deriving limit theorems and bootstrap consistency. See, for instance, [25], [10], [11], [8], [12], [18], [7], [9], [4], [19], [20], [26], [3], among many others. However, there are few results on non-asymptotic concentration bounds for V- and U-statistics.

*University of Washington, United States of America. E-mail: ydshen@uw.edu, fanghan@uw.edu, dwitten@uw.edu
Exponential inequalities for dependent V-statistics

Exceptions include [5] and [16], who proved Hoeffding-type inequalities for U- and V-statistics under φ-mixing conditions. There, the results either rely on assumptions difficult to verify, or are limited to nondegenerate ones.

In this paper, we show that for a strongly mixing stationary sequence, exponential inequalities hold for a large class of V- and U-statistics. The main theorem is presented in Section 2. We then illustrate the usefulness of our theory with examples and some further extensions in Section 3. Detailed proof of the main theorem is given in Section 4, with the rest of proofs given in Section 5.

Notation used in the rest of the paper is as follows. $L_1(\mathbb{R}^d)$ denotes the class of integrable functions in $\mathbb{R}^d$, and for each $p \geq 1$, $\|f\|_{L_p} := \left\{ \int_{\mathbb{R}^d} |f(x)|^p dx \right\}^{1/p}$. For a real vector $u \in \mathbb{R}^d$, $\|u\|$ denotes its Euclidean norm. For two real numbers $a, b$, $a \vee b := \max\{a, b\}$.

2 Main results

For two $\sigma$-algebras $\mathcal{A}$ and $\mathcal{B}$, the strong mixing coefficient is defined as

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$ 

A stationary sequence $\{X_i\}_{i \in \mathbb{Z}}$ is called strong mixing (hereafter also called $\alpha$-mixing) if

$$\alpha(i) := \alpha(M_0, G_i) \to 0 \text{ as } i \to \infty,$$

where $M_0 := \sigma(X_j, j \leq 0)$ and $G_i := \sigma(X_j, j \geq i)$ for $i \geq 1$ are the $\sigma$-algebras generated by $\{X_j, j \leq 0\}$ and $\{X_j, j \geq i\}$ respectively.

We now introduce concepts in V-statistics. Let $\tilde{X}_1^n$ be an i.i.d. sequence with $\tilde{X}_1$ identically distributed as $X_1$. The mean value of a symmetric kernel $f$ is defined as

$$\theta := \theta(f) := \mathbb{E}f(\tilde{X}_1, \ldots, \tilde{X}_m).$$

The kernel $f$ is called centered if $\theta(f) = 0$, and degenerate of level $r - 1$ ($2 \leq r \leq m$) if

$$\mathbb{E}f(x_1, \ldots, x_{r-1}, \tilde{X}_r, \ldots, \tilde{X}_m) = \theta$$

for any $(x_1, \ldots, x_{r-1})^\top \in \text{supp}(P^{r-1})$, the support of the product measure $P^{r-1}$. The kernel $f$ is called fully degenerate if it is degenerate of level $m - 1$.

When $f$ is degenerate of level $r - 1$, its Hoeffding decomposition takes the form

$$f(x_1, \ldots, x_m) - \theta = \sum_{1 \leq i_1 < \ldots < i_r \leq m} f_r(x_{i_1}, \ldots, x_{i_r}) + \ldots + f_m(x_1, \ldots, x_m),$$

where $\{f_p\}_{p=1}^m$ are recursively defined as

$$f_1(x) := g_1(x),$$

$$f_p(x_1, \ldots, x_p) := g_p(x_1, \ldots, x_p) - \sum_{k=1}^{p} f_1(x_k) - \ldots - \sum_{1 \leq k_1 < \ldots < k_{p-1} \leq p} f_{p-1}(x_{k_1}, \ldots, x_{k_{p-1}}),$$

for $p = 2, \ldots, m$, with $\{g_p\}_{p=1}^m$ defined as $g_m := f - \theta$, and

$$g_p(x_1, \ldots, x_p) := \mathbb{E}f(x_1, \ldots, x_p, \tilde{X}_{p+1}, \ldots, \tilde{X}_m) - \theta$$


http://www.imstat.org/ejp/
for $1 \leq p \leq m - 1$. For each $1 \leq p \leq m$ and $1 \leq k \leq n$, we denote the V-statistic generated by $f_p$ and data $\{X_i\}_{i=1}^k$ by
\[
V_k(f_p) := \sum_{i_1, \ldots, i_p=1}^k f_p(X_{i_1}, \ldots, X_{i_p}).
\]

For a real function $g \in L_1(\mathbb{R}^d)$, its Fourier transform is defined as
\[
\hat{g}(u) := \int_{\mathbb{R}^d} g(x)e^{-2\pi i u^\top x} \, dx,
\]
where $dx := dx_1 \ldots dx_d$.

**Theorem 2.1.** Suppose $\{X_i\}_{i=1}^n$ in (1.1) is part of a stationary sequence $\{X_i\}_{i \in \mathbb{Z}}$ that is geometrically $\alpha$-mixing with coefficient
\[
\alpha(i) \leq \gamma_1 \exp(-\gamma_2 i) \quad \text{for all } i \geq 1,
\]
where $\gamma_1, \gamma_2$ are two positive absolute constants. Suppose $f \in L_1(\mathbb{R}^m)$ is continuous, and its Fourier transform $\hat{f}$ satisfies
\[
\int_{\mathbb{R}^m} \|\hat{f}(u)\|^q \, du < \infty
\]
for some $q \geq 1$. Then, there exists a positive constant $C = C(m, \gamma_1, \gamma_2)$ such that for each $1 \leq p \leq m$, and any $x > 0$,
\[
P\left\{ n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)| \geq x \right\} \leq 6 \exp \left( -\frac{C n x^{2/p}}{A_{p,n}^{1/p} + x^{1/p} M_{p,n}^{1/p}} \right)
\]
with
\[
A_{p,n} = 2^m \|f\|_{L_1}^2 \left\{ \frac{64 \gamma_1^{1/3}}{1 - \exp(-\gamma_2/3)} + \frac{(\log n)^4}{n} \right\}^p 	ext{ and } M_{p,n} = 2^m \|\hat{f}\|_{L_1} (\log n)^{2p}.
\]

We remark that a maximal-type tail estimate for $V_n(f)$ in (1.1) can be obtained in a straightforward manner by assembling the tail estimate in (2.4) for each $r \leq p \leq m$, where $r$ is the degenerate level of $f$. Indeed, for any $1 \leq k \leq n$ and $1 \leq p \leq m$, we have by the symmetry of $f$
\[
\sum_{1 \leq i_1, \ldots, i_m \leq k} \sum_{1 \leq j_1 < \ldots < j_p \leq m} f_p(X_{i_1}, \ldots, X_{i_p}) = \binom{m}{p} \sum_{1 \leq i_1, \ldots, i_m \leq k} f_p(X_{i_1}, \ldots, X_{i_p})
\]
\[
= \binom{m}{p} k^{m-p} \sum_{1 \leq i_1, \ldots, i_p \leq k} f_p(X_{i_1}, \ldots, X_{i_p}).
\]

This entails that
\[
\begin{align*}
n^{-m} \max_{1 \leq k \leq n} & \left| \sum_{1 \leq i_1, \ldots, i_m \leq k} \left( f(X_{i_1}, \ldots, X_{i_m}) - \theta \right) \right| \\
= n^{-m} \max_{1 \leq k \leq n} & \left| \sum_{1 \leq i_1, \ldots, i_m \leq k} \sum_{1 \leq p=r}^{m} \sum_{1 \leq j_1 < \ldots < j_p \leq m} f_p(X_{i_1}, \ldots, X_{i_p}) \right| \\
\leq n^{-m} \max_{1 \leq k \leq n} & \sum_{r \leq p \leq m} \left( \binom{m}{p} k^{m-p} \right) \sum_{1 \leq i_1, \ldots, i_p \leq k} f_p(X_{i_1}, \ldots, X_{i_p}) \\
\leq & \sum_{r \leq p \leq m} \left( \binom{m}{p} n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)| \right).
\end{align*}
\]
Therefore, by adjusting the constant $C$ in (2.4), we obtain that
\[
P \left( \sum_{1 \leq k \leq n} \max_{1 \leq i, \ldots, i_m \leq k} \left| f(X_{i_1}, \ldots, X_{i_m}) - \theta \right| \geq x \right) \leq \delta \exp \left( -\frac{Cn^2x^{2/p}}{Ap_1^p + x^{1/p}M_1^{1/p}} \right).
\]

We now provide a proof sketch of Theorem 2.1 with a focus on the technical novelties. One key step in our proof is to find a uniform approximation $\tilde{f} = \tilde{f}(t, M)$ of the original kernel $f$ under any prescribed accuracy $t$ such that (i) $|\tilde{f} - f| \leq t$ uniformly over a large enough compact set $[-M, M]^{md}$, and (ii) $\tilde{f}$ admits the following tensor expansion
\[
\tilde{f}(x_1, \ldots, x_m) = \sum_{j_1, \ldots, j_m=1}^K \tilde{f}_{j_1, \ldots, j_m} e_{j_1}(x_1) \ldots e_{j_m}(x_m).
\] (2.6)

Here, $K$ is a positive integer that depends on both the approximation error $t$ and the range of approximation $M$. The elements $\tilde{f}_{j_1, \ldots, j_m}$ are a real sequence, and $\{e_j(\cdot)\}_{j=1}^K$ is a set of uniformly bounded real bases. Once such an $\tilde{f}$ is found, a truncation argument will yield the proximity between $\{f_{p, t, M}\}_{p=1}^m$ and $\{\tilde{f}_{p, t, M}\}_{p=1}^m$, the latter being the degenerate components of $f(t, M)$ in its Hoeffding decomposition. Then, using each $\tilde{f}_{p, t, M}$ as a proxy, standard moment estimates with the aid of exponential inequalities for partial sum processes (cf. Corollary 24 in [21]) will render a tail bound for each $\max_{1 \leq k \leq n} |V_k(f_{p, t, M})|$. The proof then boils down to finding such an $\tilde{f}$ with the tensor structure (2.6). One main difficulty in this step is to construct expansion bases $\{e_j(\cdot)\}_{j=1}^K$ that are uniformly bounded. Many classical approaches in multivariate function approximation are unable to provide a satisfactory answer to this problem. For example, uniform polynomial approximation by the Stone-Weierstrass theorem will have very poor performance, since high orders of the polynomials lead to a large upper bound of the bases. The use of Lipschitz-continuous scale and wavelet functions, as exploited in [19], is also inappropriate for the same reason.

Our solution is based on a probabilistic method, and especially, by realizing that the tensor decomposition (2.6) is intrinsically connected to the idea of randomized feature mapping [22] in the kernel learning literature. More specifically, when $f \in L_1(R^{md})$ is continuous and $\tilde{f} \in L_1(R^{md})$, the Fourier inversion formula implies that
\[
f(x_1, \ldots, x_m) = \int_{R^{md}} \tilde{f}(u_1, \ldots, u_m) e^{2\pi i (u_1^\top x_1 + \ldots + u_m^\top x_m)} du_1 \ldots du_m,
\]
where the right-hand side can be seen as the expectation of a Fourier basis with random frequency, which follows the sign measure of $\tilde{f}$. Due to the boundedness of the Fourier bases, Hoeffding’s inequality guarantees an exponentially fast rate for a sample mean statistic of Fourier bases
\[
s_K(x_1, \ldots, x_m) := \frac{1}{K} \sum_{j=1}^K \exp \left\{ 2\pi i (u_{j,1}^\top x_1 + \ldots + u_{j,m}^\top x_m) \right\}
\]
to approximate $f$ at each fixed point $x \in R^{md}$. The elements $\exp \{2\pi i (u_{j,1}^\top x_1 + \ldots + u_{j,m}^\top x_m)\}$ in $s_K(x_1, \ldots, x_m)$ naturally decompose to bounded basis functions of inputs $x_j$. An entropy-type argument is then used so that we could prove the existence of a satisfactory set of bases such that the approximation holds uniformly over any compact set $[-M, M]^{md}$. The detailed proof will be given in Section 4.
3 Examples and extensions

Motivated by their wide applications in statistics and machine learning, we will put special focus on shift-invariant symmetric kernels in the case \( m = 2 \) with \( f(x, y) = f_0(x - y) \) for some \( f_0 : \mathbb{R}^d \to \mathbb{R} \). We start with a corollary of Theorem 2.1 for such kernels.

**Corollary 3.1.** Let \( \{X_i\}_{i=1}^n \) be as in Theorem 2.1. Let \( m = 2 \) and the kernel \( f \) be shift-invariant with \( f(x, y) = f_0(x - y) \) for some \( f_0 : \mathbb{R}^d \to \mathbb{R} \). Suppose that \( f_0 \in L_1(\mathbb{R}^d) \) is continuous, and its Fourier transform \( \hat{f}_0 \) satisfies

\[
\int_{\mathbb{R}^d} |\hat{f}_0(u)| \|u\|^q du < \infty
\]

for some \( q \geq 1 \). Then, for \( p = 1, 2 \), the same tail bound in (2.4) holds with

\[
A_{p,n} = 16\|\hat{f}_0\|_{L_1}^2 \left\{ \frac{64\gamma_{1/3}}{1 - \exp(-\gamma_{2/3})} + \frac{(\log n)^4}{n} \right\}^p \quad \text{and} \quad M_{p,n} = 4\|\hat{f}_0\|_{L_1}(\log n)^{2p}.
\]

In view of Bochner’s theorem (cf. Section 1.4.3, [23]), Corollary 3.1 can be further simplified when the kernel is positive definite. Recall that a real function \( f \) is positive semi-definite (PSD) if and only if it is continuous and its Fourier transform is non-negative for all \( u \). A real function \( f \) is positive definite (PD) if and only if it is continuous, its Fourier transform is non-negative for all \( u \), and for any \( x \in \mathbb{R}^d \), the matrix \( A = (a_{ij})_{i,j=1}^n \) with \( a_{ij} = f_0(x_i - x_j) \) is positive semi-definite (PSD).

**Corollary 3.2.** Let \( \{X_i\}_{i=1}^n \) be as in Theorem 2.1. Let \( m = 2 \) and the kernel \( f \) be shift-invariant with \( f(x, y) = f_0(x - y) \) for some \( f_0 : \mathbb{R}^d \to \mathbb{R} \). Suppose \( f_0 \) satisfies the conditions in Corollary 3.1 and is also PD. Then, for \( p = 1, 2 \), the same tail bound in (2.4) holds with

\[
A_{p,n} = 4f_0(0)^2 \left\{ \frac{64\gamma_{1/3}}{1 - \exp(-\gamma_{2/3})} + \frac{(\log n)^4}{n} \right\}^p \quad \text{and} \quad M_{p,n} = 2f_0(0)(\log n)^{2p}.
\]

Moreover, the same bound holds with the above \( A_p \) and \( M_p \) if \( f_0 \) only satisfies (3.1) for some \( 0 < q < 1 \), but is both PD and Lipschitz continuous.

We now list several commonly-used kernels covered by Theorem 2.1 and the previous two corollaries.

1. The \( d \)-dimensional Gaussian kernel \( f(x, y) = f_0(x - y) = \exp\left(-\|x - y\|^2/2\right) \) is shift-invariant with \( f_0 \) being both Schwartz and PD, and \( f_0 \) satisfies (3.1) for arbitrary \( q \geq 1 \). Thus, \( f \) satisfies the conditions of Corollary 3.2.

2. For the \( d \)-dimensional Cauchy kernel \( f(x, y) = f_0(x - y) \) with \( f_0(x) = \prod_{i=1}^d 2/(1 + x_i^2) \), \( f_0 \) is PD and its Fourier transform \( \hat{f}_0(u) = \exp(-\|u\|_2) \) satisfies (3.1) for arbitrary \( q \geq 1 \). Therefore, it satisfies the conditions of Corollary 3.2.

3. The \( d \)-dimensional Laplacian kernel \( f(x, y) = f_0(x - y) = \exp(-\|x - y\|_1) \) is shift-invariant and PD. The Fourier transform of \( f_0 \) is \( \hat{f}_0(u) = \prod_{i=1}^d \left\{ 2/(1 + u_i^2) \right\} \), which has fractional moments and thus satisfies (3.1) for any \( 0 < q < 1 \). Since \( f_0 \) is both PD and Lipschitz, it satisfies the conditions of Corollary 3.2.

4. The 1-dimensional “hat” kernel: \( f(x, y) = f_0(x - y) \) with \( f_0(x) \) equal to \( x + 1 \) for \( -1 \leq x \leq 0 \), \( 1 - x \) for \( 0 \leq x \leq 1 \), and 0 otherwise. \( f_0 \) is PD and 1-Lipschitz. Its Fourier transform is \( \hat{f}_0(u) = (1 - \cos(2\pi u))/(2\pi^2 u^2) \) and thus has fractional moment. Therefore, \( f_0 \) satisfies (3.1) for any \( 0 < q < 1 \), and hence is also covered by Corollary 3.2.

We then discuss extensions to Theorem 2.1. The smoothness assumption (2.3) in Theorem 2.1 could be further relaxed by employing the standard smoothing technique.
through mollifiers. More precisely, we resort to an intermediate kernel $f_h$ between $f$ and $f$. It is constructed by convolving $f$ with the Gaussian mollifier with scale parameter $h$. The parameter $h$ controls the trade-off between approximation error and smoothness: small $h$ leads to finer approximation of $f$ by $f_h$, but makes $f_h$ less smooth and thus renders a larger constant in the tail bound. Theorem 2.1 is then applied on this intermediate kernel $f_h$ to obtain the tail bound.

As a particular example, the following corollary deals with Lipschitz kernels considered in [19]. Introduce the following constant from integration with polar coordinates (with convention $(-1)! = 0! = 1$):

$$
\Gamma(n) := \begin{cases} ((n - 2)!!)^{-1}(2\pi)^{\frac{n}{2}} & \text{if } n \text{ is even} \\ ((n - 2)!!)^{-1}(2\pi)^{-\frac{n}{2}} & \text{if } n \text{ is odd} \end{cases}
$$

(3.2)

**Corollary 3.3.** Let $\{X_i\}_{i=1}^n$ be as in Theorem 2.1. Suppose the kernel $f \in L_1(\mathbb{R}^{md})$ is bounded, uniformly continuous, and its Fourier transform satisfies

$$
|\hat{f}(u)| \leq \frac{L}{1 + \|u\|^{md+\varepsilon}}
$$

(3.3)

for some $\varepsilon > 0$ and positive constant $L$. Then, for $1 \leq p \leq m$, the bound in (2.4) holds with

$$
A_{p,n} = (1 + \varepsilon^{-1})^22^{m}c_1L^2\left\{ \frac{64\gamma_1^{1/3}}{1 - \exp(-\gamma_2/3)} + \frac{(\log n)^4}{n} \right\}^p,
$$

and $M_{p,n} = (1 + \varepsilon^{-1})^2m_1L(\log n)^{2p},$

where $c_1 = \Gamma(md)$.

The tail condition in (3.3) is in general milder than (2.3) in Theorem 2.1, and naturally arises in Fourier analysis (cf. Chapter 8.4 in [14]). The following is the version of Corollary 3.3 for shift-invariant kernels.

**Corollary 3.4.** Suppose $\{X_i\}_{i=1}^n$ are as in Theorem 2.1. Let $m = 2$ and the kernel $f$ be shift-invariant with $f(x,y) = f_0(x-y)$ for some $f_0 : \mathbb{R}^d \to \mathbb{R}$. Suppose that $f_0$ satisfies condition (3.3) (with $m = 1$) for some $\varepsilon > 0$ and positive constant $L$. Then, for $p = 1, 2$, the bound in (2.4) holds with

$$
A_{p,n} = 16(1 + \varepsilon^{-1})^2c_1L^2\left\{ \frac{64\gamma_1^{1/3}}{1 - \exp(-\gamma_2/3)} + \frac{(\log n)^4}{n} \right\}^p
$$

and $M_{p,n} = 4(1 + \varepsilon^{-1})c_1L(\log n)^{2p},$

where $c_1 = \Gamma(d)$.

Corollaries 3.3 and 3.4 cover the cosine kernel, defined as $f(x,y) := f_0(x-y) := \prod_{i=1}^d \cos(x_i - y_i) \mathbb{1}(|x_i - y_i| \leq \pi/2)$. Consider the simple 1-dimensional case. Here, even though the trigonometric identity $\cos(x - y) = \cos(x)\cos(y) + \sin(x)\sin(y)$ gives a direct expansion of $\cos(x-y)$, there is no trivial expansion of the indicator $\mathbb{1}(|x - y| \leq \pi/2)$. However, letting $f_0(x) = \cos(x)\mathbb{1}(|x| \leq \pi/2)$, it is immediate that $f_0$ is 1-Lipschitz and thus uniformly continuous. Moreover, its Fourier transform is $|\hat{f}_0(u)| = |2\cos(\pi^2u)/(1 - 4\pi^2u^2)|$, and hence $f_0$ satisfies (3.3) with $\varepsilon = 1$ and $L = 2$.

## 4 Proof of Theorem 2.1

We will use the following extra notation. For any real-valued function $f$ on $\mathbb{R}^d$, $\nabla_x f$ is the gradient of $f$. For a set $A$, $|A|$ indicates its cardinality. For a subset $M$ in $\mathbb{R}^d$, we will use $\text{diam}(M)$ to denote its diameter, i.e., $\text{diam}(M) := \sup_{x,y \in M} \|x - y\|$. For a function
Exponential inequalities for dependent V-statistics

We now approximate \( f \) to emphasize its dependence on some parameter \( \theta \). For a measurable set \( A \), we will use \( 1 \{ A \} \) to denote the indicator variable on the set \( A \). For any positive integer \( N \), we will use \( \{ N \} \) to denote the set \( \{1, \ldots, N\} \).

As described in the proof sketch after Theorem 2.1, we split the main part of the proof into the following lemmas. The first lemma finds a symmetric kernel \( \tilde{f} \) with tensor decomposition (2.6) that approximates \( f \) uniformly over some prescribed set \([-M, M]^m \) and accuracy \( t \).

**Lemma 4.1.** Suppose the kernel \( f \in L_1(\mathbb{R}^m) \) is continuous and satisfies condition (2.3) for some \( q \geq 1 \). Then, for any \( M > 0 \) and \( t > 0 \), there exists a symmetric function \( \tilde{f} = \tilde{f}(t, M) \) such that \( |f(x_1, \ldots, x_d) - \tilde{f}(x_1, \ldots, x_d)| \leq t \) uniformly over all \((x_1, \ldots, x_d)^T \in [-M, M]^m \), and \( \tilde{f} \) satisfies (2.6) for some positive integer \( K = K(t, M) \), \{ \tilde{f}_{j_1, \ldots, j_m} \}_{j_1, \ldots, j_m=1}^K \), and \{ \epsilon_j(\cdot) \}_{j=1}^K \) such that

\[
\sum_{j_1, \ldots, j_m=1}^K |\tilde{f}_{j_1, \ldots, j_m}| \leq F \quad \text{and} \quad \sup_{1 \leq j \leq K} \sup_{x \in \mathbb{R}^d} |\epsilon_j(x)| \leq B
\]

for some constants \( F, B \) that do not depend on \( M \) and \( t \). In particular, one can take \( F = 2^m \left\| \tilde{f} \right\|_{L_1} \) and \( B = 1 \).

**Proof.** This proof adapts from that of Claim 1 in [22]. Throughout the proof, \( x_1, \ldots, x_m \) and \( u_1, \ldots, u_m \) are real vectors in \( \mathbb{R}^d \), \( dx = dx_1 \ldots dx_d \), and \( x, u \) will be real vectors in \( \mathbb{R}^m \). Let \( \hat{f} : \mathbb{R}^m \rightarrow \mathbb{C} \) be the Fourier transform of \( f \), that is,

\[
\hat{f}(u_1, \ldots, u_m) = \int_{\mathbb{R}^m} f(x_1, \ldots, x_m) e^{-2\pi i (u_1^\top x_1 + \ldots + u_m^\top x_m)} dx_1 \ldots dx_m.
\]

Clearly, Condition (2.3) with some \( q \geq 1 \) implies that \( \hat{f} \in L_1(\mathbb{R}^m) \). Since \( f \) is continuous, by the Fourier inversion formula (see, for example, Chapter 6 of [24]), we have

\[
f(x_1, \ldots, x_m) = \int_{\mathbb{R}^d} \hat{f}(u_1, \ldots, u_m) e^{2\pi i (u_1^\top x_1 + \ldots + u_m^\top x_m)} du_1 \ldots du_m.
\]

Note that without the continuity of \( f \), the above equation only holds almost surely with respect to the Lebesgue measure. Let \( \hat{f} = \hat{g} + \hat{h} \) for real-valued functions \( \hat{g}, \hat{h} \), then since \( f \) is real-valued, we have \( f = I - II \), where

\[
I := \int_{\mathbb{R}^m} \hat{g}(u_1, \ldots, u_m) \cos \left\{ 2\pi (u_1^\top x_1 + \ldots + u_m^\top x_m) \right\} du_1 \ldots du_m,
\]

\[
II := \int_{\mathbb{R}^m} \hat{h}(u_1, \ldots, u_m) \sin \left\{ 2\pi (u_1^\top x_1 + \ldots + u_m^\top x_m) \right\} du_1 \ldots du_m.
\]

We now approximate \( I \) and \( II \) separately. \( I \) can be further written as \( I = I_+ - I_- \), where

\[
I_+ := \int_{\{u_1^\top x_1 + \ldots + u_m^\top x_m > 0\}} \hat{g}(u_1, \ldots, u_m) \cos \left\{ 2\pi (u_1^\top x_1 + \ldots + u_m^\top x_m) \right\} du_1 \ldots du_m,
\]

\[
I_- := \int_{\{u_1^\top x_1 + \ldots + u_m^\top x_m < 0\}} \hat{g}(u_1, \ldots, u_m) \cos \left\{ 2\pi (u_1^\top x_1 + \ldots + u_m^\top x_m) \right\} du_1 \ldots du_m.
\]

Let \( A_+ := \int_{\{u_1^\top x_1 + \ldots + u_m^\top x_m > 0\}} (-\hat{g}(u))du \) and \( A_- := \int_{\{u_1^\top x_1 + \ldots + u_m^\top x_m < 0\}} (-\hat{g}(u))du \), and note that \( A_+^+ \) and \( A_-^- \) are both nonnegative and satisfy \( A_+^+ + A_-^- = \| \hat{g} \|_{L_1} < \infty \) and \( A_+ - A_- = f(0) \), where we use the fact that \( \hat{g} \in L_1(\mathbb{R}^d) \) since \( f \in L_1(\mathbb{R}^md) \). Then, we have

\[
I = A_+^+ \cdot E_u \left[ \cos \left\{ 2\pi (v_1^\top x_1 + \ldots + v_m^\top x_m) \right\} \right] - A_-^- \cdot E_u \left[ \cos \left\{ 2\pi (v_1^\top x_1 + \ldots + v_m^\top x_m) \right\} \right]
\]

\[
= A_+^+ \cdot k_+^+ (x_1, \ldots, x_m) - A_-^- \cdot k_-^- (x_1, \ldots, x_m),
\]

http://www.imstat.org/ejp/
where \((u_1^T, \ldots, u_m^T)^T\) follows the distribution \(\hat{g} 1\{\hat{g} > 0\}/A_\hat{g}^+\), and \((v_1^T, \ldots, v_m^T)^T\) follows the distribution \(-\hat{g} 1\{\hat{g} < 0\}/A_\hat{g}^-\). Assume without loss of generality that \(A_\hat{g}^+ > 0\) and \(A_\hat{g}^- > 0\). We now focus on \(I_4\). For any compact subset \(\mathcal{M} \subset \mathbb{R}^{md}\), there exist \(T\) Euclidean balls with radius \(r\) that cover \(\mathcal{M}\), where \(T \leq \{c \text{diam}(\mathcal{M})/r\}^{md}\) with \(c = 3\sqrt{md}/\pi\). Denote \(\{d_1, \ldots, d_T\}\) as the centers of these balls in \(\mathbb{R}^{md}\). Now choose an i.i.d. sample \(\{(u_1^T, \ldots, u_m^T)^T\}_{i=1}^{D_1}\) from the distribution \(\hat{g} 1\{\hat{g} > 0\}/A_\hat{g}^+\) with the sample size \(D_1\) to be specified later. Then, for each center \(d = (d_1^T, \ldots, d_m^T)^T\) and any \(t > 0\), it holds by Hoeffding’s inequality that

\[
P\left(\left| \frac{1}{D_1} \sum_{i=1}^{D_1} \cos \left\{ 2\pi \left( u_i^T d_1 + \ldots + u_m^T d_m \right) \right\} - k_g^+ (d_1, \ldots, d_m) \right| \geq \frac{t}{8} \right) \leq \exp \left( -\frac{D_1 t^2}{128} \right).
\]

Let \(s_{D_1} (x_1, \ldots, x_m) := \sum_{i=1}^{D_1} \cos \left\{ 2\pi \left( u_i^T x_1 + \ldots + u_m^T x_m \right) \right\}/D_1\) so that \(k_g^+ (x_1, \ldots, x_m) = E_u \{s_{D_1} (x_1, \ldots, x_m)\}.\) Then, for any \(q \geq 1\), it holds that

\[
E \left[ \sup_x \| \nabla_x s_{D_1} (x) - k_g^+ (x) \|^q \right] = E \left[ \sup_x \| \nabla_x s_{D_1} (x) - E \nabla_x s_{D_1} (x) \|^q \right] \\
\leq E \left[ \sup_x \left\{ \| \nabla_x s_{D_1} (x) \| + E (\| \nabla_x s_{D_1} (x) \|) \right\}^q \right] \\
\leq 2^{q-1} E \left[ \sup_x \| \nabla_x s_{D_1} (x) \|^q + \sup_x \left\{ E (\| \nabla_x s_{D_1} (x) \|) \right\}^q \right] \\
\leq 2^q E \left[ \sup_x \| \nabla_x s_{D_1} (x) \|^q \right],
\]

where in the first line we use the finiteness of \(\int_{\mathbb{R}^d} |\hat{f}(u)| \|u\|^q du\) (guaranteed by Condition (2.3)) and dominated convergence theorem to exchange the derivative with expectation. Moreover,

\[
E \left( \sup_x \| \nabla_x s_{D_1} (x) \|^q \right) = E \left[ \sup_x \left\| \frac{1}{D_1} \sum_{i=1}^{D_1} 2\pi u_i \cos \left\{ 2\pi \left( u_i^T x \right) \right\} \right\|^q \right] \\
\leq (2\pi)^q E \left\{ \left( \frac{1}{D_1} \sum_{i=1}^{D_1} \| u_i \| \right)^q \right\} \\
\leq (2\pi)^q E \left( \frac{1}{D_1} \sum_{i=1}^{D_1} \| u_i \|^q \right) = (2\pi)^q E (\| u_1 \|^q),
\]

where we have used the finiteness of \(E (\| u_1 \|^q)\) since \(\int_{\mathbb{R}^d} |\hat{f}(u)| \|u\|^q du < \infty\) and the convexity of the function \(x^q\) when \(q \geq 1\). Therefore, it holds that

\[
E \left( \sup_x \| \nabla_x s_{D_1} (x) - k_g^+ (x) \|^q \right) \leq (4\pi)^q E (\| u_1 \|^q),
\]

and thus by Markov’s inequality,

\[
P \left( \sup_x \| \nabla_x s_{D_1} (x) - k_g^+ (x) \| \geq \frac{t}{8} \right) \leq \left( \frac{32\pi r}{t} \right)^q E (\| u_1 \|^q).
\]

By the triangle inequality, the event \(\sup_{x \in \mathcal{M}} \{s_{D_1} (x) - k_g^+ (x) \leq t/4\}\) has greater probability than the following event

\[
\left\{ \sup_{x \in \mathcal{M}} \| s_{D_1} (x) - k_g^+ (x) \| \leq t/8, \forall d \in \{d_1, \ldots, d_T\} \right\} \cap \left\{ \sup_x \| \nabla_x s_{D_1} (x) - k_g^+ (x) \| \leq t/(8r) \right\}
\]

Therefore, we have

\[
P \left( \sup_{x \in \mathcal{M}} \| s_{D_1} (x) - k_g^+ (x) \| \geq \frac{t}{4} \right) \leq \left( \frac{c \text{diam}(\mathcal{M})}{r} \right)^{md} \exp \left( -\frac{D_1 t^2}{128} \right) \left( \frac{32\pi r}{t} \right)^q E (\| u_1 \|^q).
\]
Letting the right-hand side of the above inequality be of the form $\kappa_1 r^{-md} + \kappa_2 r^q$, and $r = (\kappa_1/\kappa_2)^{1/(q+md)}$, we have
\[
P\left( \sup_{x \in M} \left| s_{D_1}(x) - k_y^+(x) \right| \geq \frac{t}{4} \right) \leq 2 \left( \frac{32\pi E\|u_1\|_q^1}{t} \right)^{\frac{md}{q}} \exp \left( -\frac{D_1 t^2}{128} \frac{q}{q + md} \right).
\]

Now, using the fact
\[
E\|u_1\|^q = \int_{\mathbb{R}^md} \|u\|_q \hat{g}(u) 1\{\hat{g}(u) > 0\} du \leq \frac{1}{A_y^+} \int_{\mathbb{R}^md} \|u\|_q \hat{g}(u) du \leq \frac{1}{A_y^+} \int_{\mathbb{R}^md} \|u\|_q \hat{f}(u) du,
\]
we conclude that there exists $\{u_i\}_{i=1}^{D_1} \in \mathbb{R}^md$ such that uniformly over $M$, it holds that
\[
A_y^+ \cdot \left| s_{D_1}(x) - k_y^+(x) \right| = \left| \frac{A_y^+}{D_1} \sum_{i=1}^{D_1} \cos \{2\pi (u_i^+ x)\} - A_y^+ \cdot k_y^+(x) \right| \leq \frac{A_y^+ t}{4}
\]
when $D_1$ is chosen such that
\[
D_1 \geq C_1 \frac{md}{t^2} \log \left( \frac{\pi c \text{diam}(M) t q \hat{f}(\hat{f})}{(A_y^+)^{1/q} t} \right)
\]
for some sufficiently large constant $C_1 = C_1(t, f, M)$. Equivalently, it holds that
\[
\left| A_y^+ \cdot s_{D_1}(x) - A_y^+ \cdot k_y^+(x) \right| \leq t/4
\]
when $D_1$ is chosen such that
\[
D_1 \geq C_1 \frac{md(A_y^+)^2}{t^2} \log \left( \frac{\pi c \text{diam}(M)(A_y^+)^{1-1/q} \hat{f}(\hat{f})}{t} \right).
\]

Similarly, it can be shown that there exists $\{v_i\}_{i=1}^{D_2} \in \mathbb{R}^md$ such that
\[
\left| A_y^+ \cdot s_{D_2}(x) - A_y^+ \cdot k_y^+(x) \right| \leq t/4
\]
uniformly over $x \in M$, where
\[
s_{D_2}(x) = \frac{1}{D_2} \sum_{i=1}^{D_2} \cos \{2\pi (v_i^+ x)\},
\]
and $D_2$ is chosen such that
\[
D_2 \geq C_2 \frac{md(A_y^-)^2}{t^2} \log \left( \frac{8\pi c \text{diam}(M)(A_y^-)^{1-1/q} \hat{f}(\hat{f})}{t} \right)
\]
for some sufficiently large constant $C_2 = C_2(t, f, M)$. Repeating this procedure for the approximation of $II$, then with $A_h^+, A_h^-, k_h^+, k_h^-$ similarly defined as $A_y^+, A_y^-, k_y^+, k_y^-$, we can find $s_{D_3}$ and $s_{D_4}$ which are sample means of sine functions such that
\[
\left| A_h^+ \cdot s_{D_3}(x) - A_h^+ \cdot k_h^+(x) \right| \leq t/4 \quad \text{and} \quad \left| A_h^- \cdot s_{D_4}(x) - A_h^- \cdot k_h^-(x) \right| \leq t/4
\]
uniformly over all $x \in M$, when the sample sizes $D_3$ and $D_4$ are respectively chosen such that
\[
D_3 \geq C_3 \frac{md(A_h^+)^2}{t^2} \log \left( \frac{\pi c \text{diam}(M)(A_h^+)^{1-1/q} \hat{f}(\hat{f})}{t} \right),
\]
\[
D_4 \geq C_4 \frac{md(A_h^-)^2}{t^2} \log \left( \frac{\pi c \text{diam}(M)(A_h^-)^{1-1/q} \hat{f}(\hat{f})}{t} \right)
\]
Exponential inequalities for dependent V-statistics

for some sufficiently large constants $C_3, C_4$ that depend on $t, f, M$. Putting together the pieces, we obtain that

$$|s_D(x) - f(x)| := \left| \left\{ A_y^+ \cdot s_{D_1}(x) - A_y^- \cdot s_{D_2}(x) - A_h^+ \cdot s_{D_3}(x) + A_h^- \cdot s_{D_4}(x) \right\} - f(x) \right|$$

is smaller than $t$ when $D_1-D_4$ are chosen as above. Since

$$A_y^+ + A_y^- + A_h^+ + A_h^- = \int_{\mathbb{R}^m} |\hat{g}| + |\hat{h}| \leq \sqrt{2} \int_{\mathbb{R}^m} \sqrt{|\hat{g}|^2 + |\hat{h}|^2} = 2 \left\| f \right\|_{L_1}$$

and for each $u$, $\cos \left\{ 2\pi (u^T x) \right\}$ can be written as at most $2^{m-1}$ linear combinations of the term $z_u (2\pi u^T x_1) \ldots z_u m (2\pi u^T x_m)$, where $\{z_u(\cdot)\}_{i=1}^m$ is either the cosine or sine function.

Therefore, taking $M = [-M, M]^{md}$, it holds that $|s_D - f| \leq t$ uniformly over $[-M, M]^{md}$, and $s_D$ satisfies (4.1) with constants $F = 2m \left\| f \right\|_{L_1}$ and $B = 1$. Lastly, define the symmetrized version of $s_D$ to be

$$\tilde{s}_D(x_1, \ldots, x_m) := \frac{1}{m!} \sum_\pi s_D(\pi(x_1), \ldots, \pi(x_m)),$$

where the summation is taken over all $m!$ permutations of $(x_1, \ldots, x_m)$. Then, due to the symmetry of $f$, $|\tilde{s}_D - f| \leq t$ uniformly over $[-M, M]^{md}$ and $\tilde{s}_D$ satisfies (4.1) with the same $F, B$ as $s_D$. \hfill \Box

The second lemma builds upon the previous one and guarantees the existence of an approximating kernel $\tilde{f}$ such that $f_p$ and $\tilde{f}_p$, the $p$th term in the Hoeffding decomposition of $f$ and $\tilde{f}$, are sufficiently close for each $1 \leq p \leq m$.

**Lemma 4.2.** Suppose the kernel $f \in L_1(\mathbb{R}^{md})$ is continuous and satisfies condition (2.3) for some $q \geq 1$. Then, for any $M > 0$ and $t > 0$, there exists a symmetric function $\tilde{f} = \tilde{f}(; t, M)$ such that $\tilde{f}$ satisfies all the properties in Lemma 4.1, and moreover, for each $1 \leq p \leq m$,

$$\left| f_p(x_1, \ldots, x_p) - \tilde{f}_p(x_1, \ldots, x_p) \right| \leq Ct$$

uniformly over all $(x_1^\top, \ldots, x_p^\top) \in [-M, M]^{pd}$ for some positive constant $C = C(m)$.

**Proof.** To highlight dependence, for any $t_0 > 0$ and $M_0 > 0$, we will denote the approximating kernel in Lemma 4.1 by $\tilde{f} = \tilde{f}(; t_0, M_0)$, so that $|f - \tilde{f}| \leq t_0$ uniformly over $[-M_0, M_0]^{md}$ and $\tilde{f}$ satisfies (4.1) with $F = 2m \left\| \tilde{f} \right\|_{L_1}$ and $B = 1$. This implies that

$$\sup_{(x_1, \ldots, x_d) \in \mathbb{R}^d} \left| \tilde{f}(x_1, \ldots, x_d; t_0, M_0) \right| \leq 2m \left\| \tilde{f} \right\|_{L_1}, \tag{4.3}$$

and, in particular, $\tilde{f} \in L_1(\mathbb{R}^{md})$ under the product measure $P^m$. We will prove Lemma 4.2 by choosing a $\tilde{f}(; t_0, M_0)$ with some $t_0$ and $M_0$ to be specified later that only depend on the prescribed $t$ and $M$. Again, to show clearly the dependence on $t_0$ and $M_0$, we write $\tilde{\theta}(t_0, M_0), \tilde{f}_p(; t_0, M_0),$ and $\tilde{g}_p(; t_0, M_0)$ in the Hoeffding decomposition of $\tilde{f}(; t_0, M_0)$. By definition, in order to show $|f_p - \tilde{f}_p(; t_0, M_0)| \leq Ct$ over $[-M, M]^{md}$ for some $C = C(m)$, it suffices to show that $|\theta - \tilde{\theta}(t_0, M_0)| \leq 2t_0$ and $|g_p - \tilde{g}_p(; t_0, M_0)| \leq 2t_0$ over $[-M_0, M_0]^{pd}$.
Exponential inequalities for dependent V-statistics

as long as we choose \( t_0 \leq t \) and \( M_0 \geq M \). For \( \left| \theta - \tilde{\theta}(t_0, M_0) \right| \), recalling that \( \{\tilde{X}_i\}_{i=1}^n \) are i.i.d. with law \( P \), one has

\[
\left| \theta - \tilde{\theta}(t_0, M_0) \right| = \left| E \left\{ f(\tilde{X}_1, \ldots, \tilde{X}_m) - \tilde{f}(\tilde{X}_1, \ldots, \tilde{X}_m; t_0, M_0) \right\} \right| \\
\leq E \left\{ \left| f - \tilde{f} \right| : t_0, M_0 \right\} \left\| \left( \tilde{X}_1^\top, \ldots, \tilde{X}_m^\top \right) \in [-M_0, M_0]^m \right\} + E \left\{ \left| f - \tilde{f} \right| : t_0, M_0 \right\} \left\| \left( \tilde{X}_1^\top, \ldots, \tilde{X}_m^\top \right) \notin [-M_0, M_0]^m \right\} \\
\leq t_0 + E \left\{ \left| f - \tilde{f} \right| : t_0, M_0 \right\} \left\| \left( \tilde{X}_1^\top, \ldots, \tilde{X}_m^\top \right) \notin [-M_0, M_0]^m \right\}.
\]

Now, by (4.3), the variable \( \left| f - \tilde{f} \right| \left\{ \left( \tilde{X}_1^\top, \ldots, \tilde{X}_m^\top \right) \notin [-M_0, M_0]^m \right\} \) has an integrable majorant \( |f| + 2M \left\| \tilde{f} \right\|_L_1 \) under the product measure \( P^m \), and clearly converges to zero in probability as \( M_0 \to \infty \). Thus, by choosing \( t_0 = t \) and the dominated convergence theorem, there exists some \( M_1 = M_1(t) \) such that for each \( M_0 \geq M_1(t), \left| \theta - \tilde{\theta}(t, M_0) \right| \leq 2t \). With a similar argument, there exists some \( M_2 = M_2(t) \) such that for any \( M_0 \geq M_2(t) \) and \( 1 \leq p \leq m \), it holds that

\[
|g_p(x_1, \ldots, x_p) - \tilde{g}_p(x_1, \ldots, x_p; t, M_0)| \leq 2t, \quad \text{for all } (x_1^\top, \ldots, x_p^\top) \in [-M_0, M_0]^p.
\]

Therefore, by choosing \( M_0 := M \lor M_1(t) \lor M_2(t) \), one has \( \tilde{f}(t, M) \) satisfies all the desired properties. This completes the proof. \( \square \)

The third lemma derives a maximal-type tail bound for each \( \max_{1 \leq k \leq n} |V_k(f_p)| \) when \( f \) admits the tensor decomposition (2.6).

**Lemma 4.3.** Suppose \( \{X_i\}_{i=1}^n \) are as in Theorem 2.1. Suppose the symmetric kernel \( f : R^md \to R \) can be written as

\[
f(x_1, \ldots, x_m) = \sum_{j_1, \ldots, j_m = 1}^K f_{j_1, \ldots, j_m} e_{j_1}(x_1) \cdots e_{j_m}(x_m),
\]

where \( K \) is some positive integer, \( \{f_{j_1, \ldots, j_m} \}_{j_1, \ldots, j_m = 1}^K \) is a real sequence, and \( \{e_j(\cdot)\}_{j=1}^K \) is a set of real basis functions satisfying

\[
\sum_{j_1, \ldots, j_m = 1}^K |f_{j_1, \ldots, j_m}| \leq F \quad \text{and} \quad \sup_{1 \leq j \leq K} \sup_{x \in R^d} |e_j(x)| \leq B
\]

for some positive constants \( F \) and \( B \). Let \( \mu_o := \sup_{1 \leq j \leq K} (E|e_j(X_1)|^a)^{1/a} \) for each \( a \geq 1 \). Then, there exists a positive constant \( C = C(m, \gamma_1, \gamma_2) \) such that for any \( 1 \leq p \leq m \), and any \( x \geq 0 \),

\[
P \left( n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)| \geq x \right) \leq 6 \exp \left( -\frac{C n^{1/a} 2^{p/m} \mu_o^{1/p} \gamma_2^{2/p} \log n^{2/p}}{A_{p,n}^{1/p} + x^{1/p} M_{p,n}^{1/p}} \right),
\]

where

\[
A_{p,n} = \mu_2^{2(n-p)} F^2 (\gamma_1^2 + B^2 (\log n)^4 / n)^p, \quad M_{p,n} = \mu_3^{n-p} F B^p (\log n)^2 n,
\]

and \( \sigma^2 = 64 \gamma_1^3 / (1 - \exp(-\gamma_2/3)) \).
Exponential inequalities for dependent V-statistics

Proof. Throughout the proof, let $C_i$’s be positive constants that only depend on $m, \gamma_1, \gamma_2$, and we will use the shorthand $f_{j_a,b}$ for $f_{j_a,...,j_b}$ for positive integers $a < b$. We drop the dependence of $A_{p,n}$ and $M_{p,n}$ on $n$ for notational simplicity.

Fix a $1 \leq p \leq m$ and we now derive the tail bound for $\max_{1 \leq k \leq n} |V_k(f_p)|$. For the set of bases $\{e_j(\cdot)\}_{j=1}^K$ in the expansion of $f$, define $\tilde{e}_j := e_j - E[e_j(X_1)]$ for $j \in [K]$. Since $f$ is symmetric, for any $(x_1, \ldots, x_m) \in \mathbb{R}^{md}$, $f(x_1, \ldots, x_m) = f(\pi(x_1), \ldots, \pi(x_m))$ for any permutation $\pi$ of $\{x_1, \ldots, x_m\}$. By the definition of $\{f_p\}_{p=1}^m$ in (2.1), one can readily check that

$$f_p(x_1, \ldots, x_p) = \sum_{j_1, \ldots, j_m=1}^K f_{j_1,\ldots,j_m} E(e_{j_1}) \ldots E(e_{j_{m-p}}) \tilde{e}_{j_{m-p+1}}(x_1) \ldots \tilde{e}_{j_m}(x_p),$$

for $r \leq p \leq m$. Thus, we have

$$V_k(f_p) = \sum_{j_1, \ldots, j_m=1}^K f_{j_1,\ldots,j_m} E(e_{j_1}) \ldots E(e_{j_{m-p}}) \left\{ \sum_{i=1}^k \tilde{e}_{j_{m-p+1}}(X_i) \right\} \ldots \left\{ \sum_{i=1}^k \tilde{e}_{j_m}(X_i) \right\}.$$

Define, for each $j \in [K]$ and $k \in [n]$,

$$S_{k,j} := \sum_{i=1}^k \tilde{e}_j(X_i) \quad \text{and} \quad Z_j := \max_{1 \leq k \leq n} |S_{k,j}|.$$

Note that for each $j \in [K]$, $\{\tilde{e}_j(X_i)\}_{i=1}^n$ is also geometrically $\alpha$-mixing. We now control each even order moment of $\max_{1 \leq k \leq n} |V_k(f_p)|$. Let

$$T_p := n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)|.$$

Define

$$\nu_j = C_3(n\sigma_j^2 + B^2), \quad \nu = C_3(n\sigma^2 + B^2), \quad c = C_4B(\log n)^2$$

and $\sigma^2 = \sup_{j \in [K]} \sigma_j^2$, with

$$\sigma_j^2 := \text{Var}[\tilde{e}_j(X_1)] + 2 \sum_{i>1} |\text{Cov}[\tilde{e}_j(X_1), \tilde{e}_j(X_i)]|.$$

Integrating the tail estimate in Corollary 24 of [21] and using Theorem 2.3 in [6] yield that, for any positive integer $N$, by choosing $C_4$ in $c$ to be sufficiently large,

$$E(Z_j^{2pN}) = (4pN) \int_0^\infty x^{2pN-1} P(Z_j \geq x) dx$$

$$= 4pN \cdot \left( \int_0^{CB \log n} x^{2pN-1} P(Z_j \geq x) dx + \int_{CB \log n}^\infty x^{2pN-1} P(Z_j \geq x) dx \right)$$

$$\leq C(B \log n)^{2pN} + (pN)!(8\nu)^N + (2pN)!(4c)^{2pN}$$

$$\leq (pN)!(8\nu)^N + (2pN)!(5c)^{2pN}.$$

Then, employing a similar argument as in [5] (cf. Equation (12) therein), it holds that
\[
ET_p^{2N} = \min_{1 \leq k_1, \ldots, k_{2N} \leq n} \sum_{j_1, \ldots, j_{2N} = 1}^{K} f_{j_1 \cdot \cdots \cdot f_{j_{(2N-1)m+1:2mN}} \cdot E(e_{j_1}) \cdots E(e_{j_{2N-1}}} \cdots \cdot f_{j_{(2N-1)m+1:2mN}} \cdot E(e_{j_{2N-1:2mN}}) \cdots 
\]
\[
\leq n^{-2N} \mu_1^{2N(m-p)} \sum_{j_1, \ldots, j_{2N} = 1}^{K} |f_{j_1 \cdot \cdots \cdot f_{j_{(2N-1)m+1:2mN}}} | |f_{j_{(2N-1)m+1:2mN}} | (E|Z_{j_{m+1:2mN}} |) \cdot \cdots \cdot (E|Z_{j_{2mN}} |) \}
\]
\[
\leq n^{-2N} \mu_1^{2N(m-p)} \sum_{j_1, \ldots, j_{2N} = 1}^{K} |f_{j_1 \cdot \cdots \cdot f_{j_{(2N-1)m+1:2mN}}} | (E|Z_{j_{m+1:2mN}} ^{2N} \mu_1^{\frac{1}{2N}} \cdots \cdot (E|Z_{j_{2mN}} ^{2N} \mu_1^{\frac{1}{2N}} \cdots 
\]
\[
\leq n^{-2N} \mu_1^{2N(m-p)} F^{2N} \{ (pN)! (8p)^p N + (2pN)! (4e)^{2pN} \},
\]
where in the second inequality we use the generalized Hölder inequality. By Stirling’s approximation formula \(\sqrt{2\pi} n^{n+1/2} e^{-n} \leq n! \leq e^{n+1/2} e^{-n}\), it holds that
\[
\{ (pN)! \}^{1/p} \leq e^{1/p} (pN)^{\frac{1}{2}+1/2} e^{-N} \leq C^0 N^{N+1/2} e^{-N} \leq C^0 N!.
\]

Similarly, we have \(\{ (2pN)! \}^{1/p} \leq C^0_n (2N)!\). Thus we have
\[
E(T_p) \leq \left( \frac{N}{p} \right)^{2N} \{ C^0_n N! \nu^N + C^0 (2N)! \nu^{2N} \}.
\]

Now we control the Laplace transform of \(T_p\),
\[
E(e^{\lambda T_p / p}) = \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} E(T_p^{n/p}) \leq \sum_{N=0}^{\infty} \frac{\lambda^{2N}}{(2N)!} E(T_p^{2N/p})
\]
\[
\leq 3 \left\{ \sum_{N=0}^{\infty} \frac{\lambda^{2N}}{(2N)!} C^0_n n^{-2N} \mu_1^{2N(m-p) / p} F^{2N/p} \nu^N + \sum_{N=0}^{\infty} \lambda^{2N} n^{-2N} C^0_n C^0_8 \mu_1^{2N(m-p) / p} F^{2N/p} \nu^{2N} \right\},
\]
(4.4)

where in the first inequality we use only the even moments with an absolute constant 3.

For the first summand in (4.4), we have
\[
\sum_{N=0}^{\infty} \lambda^{2N} n^{-2N} C^0_n \mu_1^{2N(m-p) / p} F^{2N/p} \nu^{2N} \leq \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} 2^{-N} C^0_n \mu_1^{2N(n-p) / p} F^{2N/p} \nu^{2N} \mu N
\]
\[
= \exp \left\{ C_1 (\lambda/n)^2 \mu_1^{2N(m-p) / p} F^{2N/p} \nu \right\},
\]

where in the first line we use the relation \(N!/(2N)! \leq 2^{-N} N!\). For the second summand, we have
\[
\sum_{N=0}^{\infty} \lambda^{2N} n^{-2N} C^0_n \mu_1^{2N(m-p) / p} F^{2N/p} \nu^{2N} \leq 1 + \frac{\lambda^N}{N!} 2 C^0_n \mu_1^{2N(m-p) / p} F^{2N/p} \nu^{2N} \mu N
\]
\[
\leq 1 + \frac{\lambda^N}{N!} 2 C^0_n \mu_1^{2N(m-p) / p} F^{2N/p} \nu^{2N} \mu N
\]

for \(\lambda \leq n / \left\{ C_8 \mu_1^{(m-p) / p} F_{1/p} \right\} \). Now using the relation \(e^x + 1 + y \leq 2e^{x+y} \) which holds for all positive \(x, y\), we have
\[
E\left(e^{\lambda T_p / p}\right) \leq 6 \exp \left[ \frac{(\lambda/n)^2 C_{11} \mu_1^{2N(m-p) / p} F^{2N/p} \nu + \lambda^N}{2 \left(1 - (\lambda/n) C_8 \mu_1^{(m-p) / p} F_{1/p}\right)} \right] = 6 \exp \left( \frac{(\lambda/n)^2 C_{11} n A_{1/p}}{2 \left(1 - (\lambda/n) C_8 n A_{1/p}\right)} \right).
\]
Exponential inequalities for dependent V-statistics

Now, taking \( \lambda = nx^{1/p}/(C_{11nA^{1/p}} + C_{12M^{1/p}}x^{1/p}) \) in the exponential Markov inequality, we have

\[
P(T_p \geq x) \leq \exp(-\lambda x^{1/p})E\left(e^{\lambda T_p^{1/p}}\right) \leq 6 \exp\left(-\frac{C_{12nA^{2/p}}}{A_p^{1/p} + x^{1/p}M_p^{1/p}}\right).
\]

Moreover, taking \( \delta = 1 \) in Theorem 3 of [13], we obtain

\[
\text{Var}\{\tilde{e}_j(X_1)\} + 2 \sum_{i \geq 1} |\text{Cov}\{\tilde{e}_j(X_1), \tilde{e}_j(X_i)\}| \leq 2 \sum_{i \geq 1} |\text{Cov}\{\tilde{e}_j(X_1), \tilde{e}_j(X_i)\}|
\]

\[
\leq 2 \left\{ \sum_{n=0}^{\infty} 8n^{1/3}(n) \right\} \|\tilde{e}_j(X_1)\|_3\|\tilde{e}_j(X_i)\|_3 \leq 64 \left\{ \sum_{n=0}^{\infty} \alpha^{1/3}(n) \right\} \|e_j(X_1)\|_3\|e_j(X_i)\|_3
\]

\[
\leq 64\gamma_1^{1/3}\mu_3^2 \sum_{n=0}^{\infty} \exp\{-(\gamma_2n/3)\} = \frac{64\gamma_1^{1/3}}{1 - \exp\{-(\gamma_2/3)\}}\mu_3^2.
\]

Putting together the pieces completes the proof. \( \square \)

We now use Lemmas 4.1–4.3 to complete the proof of Theorem 2.1.

**Proof of Theorem 2.1.** Fix a \( 1 \leq p \leq m \). Fix some \( t > 0 \) and \( M > 0 \), and define the event

\[
E := \{ X_i \in [-M, M]^d \text{ for all } 1 \leq i \leq n \}.
\]

Then, for the prescribed \( t \) and \( M \), Lemma 4.2 implies that there exists a symmetric kernel \( \tilde{f} = \tilde{f}(; t, M) \) such that \( |f - \tilde{f}(; t, M)| \leq t \) uniformly over all \( (x_1^\top, \ldots, x_m^\top)^\top \in [-M, M]^md \), and for each \( 1 \leq p \leq m \), \( |f_p - \tilde{f}_p(; t, M)| \leq Ct \) uniformly over \( [-M, M]^pd \) for some \( C = C(m) \). By definition, this implies that \( n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)| - \max_{1 \leq k \leq n} |V_k(\tilde{f}_p)| \leq Ct \) on the event \( E \), and thus for any \( x > 0 \),

\[
P(n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)| \geq x + Ct)
\]

\[
= P\left(\left\{ n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)| \geq x + Ct \right\} \cap E\right) + P\left(\left\{ n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)| \geq x + Ct \right\} \cap E^c\right)
\]

\[
\leq P\left( n^{-p} \max_{1 \leq k \leq n} |V_k(\tilde{f}_p(; t, M))| \geq x \right) + nP(X_1 \notin [-M, M]^d).
\]

Again by Lemma 4.2, \( \tilde{f}(; t, M) \) satisfies the conditions of Lemma 4.3 with constants \( F = 2^m \| \tilde{f} \|_{L_1} \) and \( B = 1 \). Therefore, applying the trivial bound that \( \mu_3 \leq B = 1 \) in Lemma 4.3, we obtain that

\[
P\left( n^{-p} \max_{1 \leq k \leq n} |V_k(f_p)| \geq x + Ct \right) \leq 6 \exp\left(-\frac{C_{12nA^{2/p}}}{A_p^{1/p} + x^{1/p}M_p^{1/p}}\right) + nP(X_1 \notin [-M, M]^d),
\]

where \( A_p \) and \( M_p \) are defined in (2.5). Now, note that the first summand on the right hand side does not depend on \( M \) or \( t \). Accordingly, by first choosing a large enough \( M \) that depends only on \( x, n, F \), since the measure \( P \) considered in this paper is always tight, we obtain that the second term is smaller than an arbitrary small proportion of the first term. Lastly, choosing \( t = x \) and adjusting the constant finishes the proof. \( \square \)
5 Proofs of other results

We will only prove Corollaries 3.1–3.3. The proof of Corollary 3.4 is similar to that of Corollary 3.1.

Proof of Corollary 3.1. By inspection of the proof of Theorem 2.1, it suffices to prove that when \( f_0 \) satisfies (3.1) for some \( q \geq 1 \), the conclusion of Lemma 4.1 still holds with \( F = 4 \| \hat{f}_0 \|_{L_1} \) and \( B = 1 \). Now, following the proof of Lemma 4.1 with prescribed range \( 2M \) and approximation error \( t \), there exists an \( \tilde{f}_0 \) with expansion in the cosine bases \( \{ \cos (2\pi u^\top x) \} \) such that \( |f_0 - \tilde{f}_0| \leq t \) uniformly over \([-2M, 2M]^d \) and \( \tilde{f}_0 \) satisfies the (4.1) with constants \( F = 2 \| \tilde{f}_0 \|_{L_1} \) and \( B = 1 \). Let \( \tilde{f}(x, y) := \tilde{f}_0(x - y) \). Then, \( |f - \tilde{f}| \leq t \) uniformly over \([-M, M]^d \), and by the trigonometric identity
\[
\cos \{ 2\pi u^\top (x - y) \} = \cos (2\pi u^\top x) \cos (2\pi u^\top y) + \sin (2\pi u^\top x) \sin (2\pi u^\top y),
\]
\( \tilde{f} \) satisfies (4.1) with constants \( F = 4 \| \hat{f}_0 \|_{L_1} \) and \( B = 1 \). This completes the proof. \( \square \)

Proof of Corollary 3.2. Again, we only need to reprove Lemma 4.1. When \( f_0 \) is PD, we have by definition that \( f_0(0) \geq 0 \) and for each \( x, y \in \mathbb{R}^d \), \( f_0(x - y) = f_0(y - x) \) implies that \( f_0(x) = f_0(-x) \) for any \( x \in \mathbb{R}^d \). This implies that the Fourier transform \( \hat{f}_0 \) of \( f_0 \) is real-valued, and \( \hat{h} = 0 \) in the proof of Lemma 4.1. Moreover, since \( f_0 \in L_1(\mathbb{R}^d) \), it satisfies (3.1) for some \( q \geq 1 \), \( f \) equals the inverse Fourier transform of \( \tilde{f} \). Thus, by Bohr’s theorem (cf. Section 1.4.3, [23]), \( \hat{f}_0 \) is nonnegative and we have \( f_0 = I = I_+ \) with \( m = 1 \) in the proof of Lemma 4.1. By definition, we have
\[
f_0(x) = \int_{\mathbb{R}^d} \hat{f}_0(u)e^{2\pi ix^\top u} du = \int_{\mathbb{R}^d} \hat{f}_0(u) e^{2\pi ix^\top u} du.
\]
Letting \( x = 0 \) in the above equation, we obtain \( \| \hat{f}_0 \|_{L_1} = f_0(0) \).

Now consider the case where \( \tilde{f}_0 \) only has fractional moment. Let \( \tilde{f}_0 := f_0/f_0(0) \) and denote the Lipschitz constants of \( \tilde{f}_0 \) and \( f_0 \) as \( L_{\tilde{f}_0} \) and \( L_{f_0} \), respectively. Then, \( L_{\tilde{f}_0} = L_{f_0}/f_0(0) \). Now we proceed with the proof of Lemma 4.1 until (4.2), and replace it with
\[
E \left\{ \sup_x \| \nabla x s_D(x) - \tilde{f}_0(x) \|^{q} \right\} \leq E \left\{ \left( \sup_x \| \nabla x s_D(x) \|^{q} + \| \nabla x \tilde{f}_0(x) \|^{q} \right) \right\} \\
\leq E \left\{ \sup_x \| \nabla x s_D(x) \|^{q} \right\} + \sup_x \| \nabla x \tilde{f}_0(x) \|^{q} \\
\leq E \left\{ \sup_x \| \nabla x s_D(x) \|^{q} \right\} + L_{\tilde{f}_0}^{q},
\]
where \( s_D(x) = \sum_{i=1}^{D} \cos (2\pi u_i^\top x)/D \) (here we use the notation \( s_D \) instead of \( s_{D_i} \), since in the PD case we only need to approximate the term \( I_+ \) as argued in the first part of the corollary). Note that the original (4.2) in the proof of Lemma 4.1 no longer holds as mere fractional moment does not guarantee the exchange of derivative and expectation in its first step. For the first term in the above inequality, we have
\[
E \left\{ \sup_x \| \nabla x s_D(x) \|^{q} \right\} = E \left[ \sup_x \left\| \frac{1}{D} \sum_{i=1}^{D} 2\pi u_i \cos \{ 2\pi (u_i^\top x) \} \right\|^{q} \right] \\
\leq (2\pi)^q E \left\{ \left\| \frac{1}{D} \sum_{i=1}^{D} u_i \right\|^{q} \right\} \\
\leq (2\pi)^q E \left\{ \left\| \sum_{i=1}^{D} u_i \right\|^{q} \right\} = (2\pi)^q D^{-q} E(\|u_1\|^{q}).
\]
Exponential inequalities for dependent V-statistics

Therefore, it holds that

$$E\left\{ \sup_x \left\| \nabla_x s_D(x) - \tilde{f}_0(x) \right\|^q \right\} \leq (2\pi)^q D^{1-q} \mathbb{E}(|u|^q) + L_f^q \tilde{f}_0.$$  

Markov inequality now gives

$$\Pr\left\{ \sup_x \left\| \nabla_x (s_D(x) - \tilde{f}_0(x)) \right\| \geq \frac{t}{2r} \right\} \leq \left( \frac{2r}{t} \right)^q \left\{ (2\pi)^q D^{1-q} \mathbb{E}(|u|^q) + L_f^q \tilde{f}_0 \right\}.$$  

Proceeding with the proof of Lemma 4.1, we obtain

$$\Pr\left( \sup_{x \in M} \left| s_D(x) - \tilde{f}_0 \right| \geq t \right) \leq \left( \frac{2r}{t} \right)^q \left\{ (2\pi)^q D^{1-q} \mathbb{E}(|u|^q) + L_f^q \tilde{f}_0 \right\} + \left( \frac{c \text{diam}(M)}{r} \right)^{md} \exp \left( -\frac{Dt^2}{8} \right).$$  

Writing the right-hand side of the above inequality in the form $\kappa_1 r^{-md} + \kappa_2 r^q$ and letting $r = (\kappa_1/\kappa_2)^{1/(q+md)}$, we obtain

$$\Pr\left( \sup_{x \in M} \left| s_D(x) - \tilde{f}_0 \right| \geq t \right) \leq 2 \left( \frac{2c \text{diam}(M)}{t} \right)^{md} \left\{ (2\pi)^q D^{1-q} \mathbb{E}(|u|^q) + L_f^q \tilde{f}_0 \right\} \exp \left( -\frac{Dt^2}{8} \right) + \frac{q}{q + md}.$$  

For any $t > 0$, we can choose large enough $D = D(t)$ such that the right-hand side of the above inequality is arbitrarily small. The proof is complete. □

**Proof of Corollary 3.3.** Let $K(\cdot) : \mathbb{R}^{md} \to \mathbb{R}$ be the standard $md$-variate Gaussian density defined as $K(x) := \exp(-\|x\|^2/2) (2\pi)^{-md/2}$, and $K_h(x) = K(x/h) h^{-md}$ for some positive constant $h$. Define $f_h(x) := (f * K_h)(x)$. Then, it holds that

$$|f_h(x) - f(x)| = \left| \int_{\mathbb{R}^{md}} (2\pi)^{-md/2} \exp \left( -\frac{\|y\|^2}{2} \right) \{f(x - hy) - f(x)\} dy \right| \leq \int_{\mathbb{R}^{md}} (2\pi)^{-md/2} \exp \left( -\frac{\|y\|^2}{2} \right) |f(x - hy) - f(x)| dy.$$  

Denote the upper bound of $f$ as $M_f$. Then, for any $t > 0$, there exists some positive constant $A = A(M_f, m, d, t)$ such that

$$\int_{[-A, A]^{md}} (2\pi)^{-md/2} \exp \left( -\frac{\|y\|^2}{2} \right) |f(x - hy) - f(x)| dy \leq t/4.$$  

Inside $[-A, A]^{md}$, using the uniform continuity of $f$, there exists some $h = h(M_f, m, d, t)$, such that

$$\int_{[-A, A]^{md}} (2\pi)^{-md/2} \exp \left( -\frac{\|y\|^2}{2} \right) |f(x - hy) - f(x)| dy \leq t/4.$$  

Putting together the pieces, it holds that for any $t > 0$, there exists some $h = h(M_f, m, d, t)$ such that $\|f_h - f\|_\infty \leq t/2$. Since both $f$ and $K_h$ belong to $L_1(\mathbb{R}^{md})$, their Fourier transforms exist. It can be readily checked that $\mathcal{K}_h(u) = \exp(-2\pi^2 h^2 \|u\|^2)$, and thus

$$\tilde{f}_h(u) = \tilde{f}(u) \cdot \mathcal{K}_h(u) = \tilde{f}(u) \exp(-2\pi^2 h^2 \|u\|^2).$$
Using the relation \( \|f * g\|_{L_q} \leq \|f\|_{L_q} \|g\|_{L_1} \) for any \( q \geq 1 \) and \( f \in L_q(\mathbb{R}^m), g \in L_1(\mathbb{R}^m) \) and the fact that \( K_h \in L_1(\mathbb{R}^m) \), it holds that \( f_h \in L_1(\mathbb{R}^m) \). Moreover, it can readily checked that

\[
\mu_q^h(\tilde{f}_h) = \int_{\mathbb{R}^d} \tilde{f}_h(u) \|u\|^q du = \int_{\mathbb{R}^d} \tilde{f}(u) \|u\|^q \exp(-2\pi^2 h^2 \|u\|^2) du < \infty
\]

for any \( q \geq 1 \). Therefore, by Lemma 4.1, for any given \( M > 0 \) and \( t > 0 \), we can find an approximating kernel \( \tilde{f}_h = \tilde{f}_h(t) \) such that \( \|\tilde{f}_h - f_h\| \leq t/2 \) uniformly over \([-M, M]^m\), and \( \tilde{f}_h \) further satisfies (4.1) with constants \( F = 2^m \|\tilde{f}_h\|_{L_1} \) and \( B = 1 \). Choosing \( h = h(M_f, m, d, t/2) \), by the triangle inequality, we have

\[
|f - \tilde{f}_h| \leq t/2 + t/2 = t
\]

uniformly over \([-M, M]^m\). Lastly, we upper bound the term \( \|\tilde{f}_h\|_{L_1} \). To this end, we have

\[
\|\tilde{f}_h\|_{L_1} = \int_{\mathbb{R}^d} \tilde{f}(u) \exp(-2\pi^2 h^2 \|u\|^2) du \leq L \int_{\mathbb{R}^d} \frac{1}{1 + \|u\|^{md+\varepsilon}} \exp(-2\pi^2 h^2 \|u\|^2) du.
\]

Using polar coordinates, it holds that

\[
\|\tilde{f}_h\|_{L_1} \leq \Gamma(md) L \int_0^\infty \frac{r^{md-1}}{1 + r^{md+\varepsilon}} \exp(-2\pi^2 h^2 r^2) dr
\]

\[
\leq \Gamma(md) L \left( 1 + \int_1^\infty \frac{1}{r^{1+\varepsilon}} dr \right)
\]

\[
= (1 + e^{-1}) \Gamma(md) L.
\]

This completes the proof. \( \square \)

References


Exponential inequalities for dependent V-statistics


http://www.imstat.org/ejp/