Z. Wahrscheinlichkeitstheorie verw. Gebiete
67, 139-167 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Invariance Principles for von Mises and U-Statistics*

Herold Dehling[1], Manfred Denker[1] and Walter Philipp[2]

[1] Institut für Mathematische Stochastik, Universität Göttingen,
Lotzestr. 13, D-3400 Göttingen, Bundesrepublik Deutschland
[2] Department of Mathematics, University of Illinois,
1409 West Green Street, Urbana, IL 61801, USA

Summary. The almost sure approximation of von Mises-statistics and U-statistics by appropriate stochastic integrals with respect to Kiefer processes is obtained. In general these integrals are non-Gaussian processes. As applications we get almost sure versions for the estimator of the variance and for the χ^2-test of goodness of fit.

1. Introduction

Let $\{X_j, j \geqq 1\}$ be a sequence of independent identically distributed random variables with common distribution function F. Let F_n be the empirical distribution function of a sample of size n. The empirical process R is defined as

$$R(s,t) = t(F_{[t]}(s) - F(s)), \qquad s \in \mathbb{R},\ t \geqq 0$$

where $[t]$ denotes the largest integer not exceeding t. Let h: $\mathbb{R}^2 \to \mathbb{R}$ be a measurable function and let $p \geqq 1$. As in [8] we define

$$\|h\|_p = \left(\int\int |h(x,y)|^p dF(x) dF(y)\right)^{1/p} + \left(\int |h(x,x)|^p dF(x)\right)^{1/p}. \tag{1.1}$$

If $\|h\|_1 < \infty$ then the stochastic double integral

$$V_n(h) = \int\int h(x,y) R(dx,n) R(dy,n) \tag{1.2}$$

is well defined and is called a von Mises statistic. Disregarding normalizing constants and the usual symmetry assumption on h we define the U-statistic

$$U_n(h) = \sum_{1 \leqq i \neq j \leqq n} h(X_i, X_j). \tag{1.3}$$

It is closely related to the von Mises statistic. (See (1.12) below.)

---

* This work was done while the last author was a visiting professor at the Institut für Mathematische Stochastik at the University of Göttingen during the Spring of 1982. He thanks the Institut and its members for their hospitality

A separable Gaussian process $\{K(s,t), 0 \leqq s \leqq 1, t \geqq 0\}$ is called a standard Kiefer process if $K(s,0) \equiv 0$, $K(0,t) = K(1,t) \equiv 0$ and

$$EK(s,t)=0, \quad 0 \leqq s \leqq 1, \quad t \geqq 0,$$
$$EK(s,t)K(s',t') = (t \wedge t')s(1-s'), \quad 0 \leqq s \leqq s' \leqq 1, \quad t,t' \geqq 0.$$

If n is an integer then $K(s,n)$ can be best written as

$$K(s,n) = \sum_{j \leqq n} B_j(s) \tag{1.4}$$

where $\{B_j(\cdot), j \geqq 1\}$ is a sequence of independent standard Brownian bridges considered as $C[0,1]$-valued random variables.

For $\|h\|_2 < \infty$ the stochastic double integral $\int\int h(x,y)K(dx,t)K(dy,t')$, $t,t' \in \mathbb{R}$ has been defined and investigated in [4]. In much of the present paper, however, we only need these integrals for $t,t' \in \mathbb{N}$ and by the above remark these can be reduced to double integrals with respect to Brownian bridges. These latter integrals have already been defined in [8].

Our theorems show that

$$W_t(h) = \int\int h(x,y)K(dx,t)K(dy,t), \quad t \geqq 0 \tag{1.5}$$

is the canonical process to approximate the von Mises statistic in the sense that it plays the same role as Brownian motion does for the approximation of partial sums of random variables or the extremal process does for the approximation of partial maxima of random variables.

We now state our results.

Theorem 1. *Let $\{X_j, j \geqq 1\}$ be a sequence of independent random variables with common distribution function F. Let $h: \mathbb{R}^2 \to \mathbb{R}$ be a measurable function with $\|h\|_2 < \infty$. Then without changing the law of the sequence $\{X_j, j \geqq 1\}$ we can redefine it on a new probability space on which there exists a standard Kiefer process $\{K(s,t), 0 \leqq s \leqq 1, t \geqq 0\}$ such that*

$$n^{-1} \max_{m \leqq n} |V_m(h) - W_m(h^*)| \to 0 \quad \textit{in probability.} \tag{1.6}$$

Here h^* is defined by

$$h^*(x,y) = h(F^{-1}(x), F^{-1}(y)) \qquad x,y \in \mathbb{R}.$$

Theorem 2. *Let $\{X_j, j \geqq 1\}$ and h be as in Theorem 1, but instead of $\|h\|_2 < \infty$ we assume that*

$$\int\int (h(x,y)\log|h(x,y)|)^2 dF(x)dF(y) + \int (h(x,x)\log|h(x,x)|)^2 dF(x) < \infty. \tag{1.7}$$

Then the conclusion of Theorem 1 remains valid but with (1.6) replaced by

$$V_n(h) - W_n(h^*) = o(n \log\log n) \quad \textit{a.s.} \tag{1.8}$$

Theorem 3. *Let K be a standard Kiefer process and let $\|h\|_2 < \infty$ where the norm $\|\cdot\|_2$ is defined in (1.1) with respect to the uniform distribution on $[0,1]$. Then* *there exists a constant $C(h)$ depending only on h such that with probability* 1

$$\limsup_{n\to\infty} (2n\log\log n)^{-1}|W_n(h)| = C(h).$$

Corollary 1. *Under the hypotheses of Theorem* 2 *we have with probability* 1

$$\limsup_{n\to\infty} (2n\log\log n)^{-1}|V_n(h)| = C(h^*)$$

where h^* is defined in Theorem 1.

Remark. In Sect. 3 we determine $C(h)$ as the maximal eigenvalue of a certain integral operator. For certain kernels h and their statistics we shall give the numerical value $C(h)$ in Sect. 8. In particular, a recent result of Csáki [3] follows.

In most applications h has some smoothness properties. The following theorem takes care of them.

Theorem 4. *Let $\{X_j, j \geqq 1\}$ and h be as in Theorem* 1. *Suppose that in addition h has the following properties*

(1.9) $$\|h\|_{2+\delta} < \infty \quad \text{for some } \delta > 0.$$

(1.10) *There is a refining sequence of partitions $\{\alpha(r), r \geqq 1\}$ of $\mathbb{R}$, $\alpha(r) = \{A_{ir}, 1 \leqq i \leqq 2^r\}$ and $h(i,j,r) \in \mathbb{R}$, $1 \leqq i, j \leqq 2^r$ and constants C and $\gamma > 0$ such that*

$$\|h - \sum_{1\leqq i,j\leqq 2^r} h(i,j,r) 1_{A_{ir}\times A_{jr}}\|_2 \leqq C 2^{-r\gamma}.$$

Then (1.8) *holds with an error term $\ll n^{1-\lambda}$ where*

(1.11) $$\lambda = \delta/(4\alpha(2+\delta)), \qquad \alpha = (96/\gamma) + (36/\delta) + 20.$$

Theorems 1, 2 and 4 immediately yield corresponding results for the U-statistic via the relation

(1.12) $$\begin{aligned} U_n(h) - n(n-1)c_1 = {} & V_n(h) + n(c_1 - c_2) - \sum_{j\leqq n} (h(X_j, X_j) - c_2) \\ & + n \sum_{j\leqq n} \left(\int h(x, X_j)\,dF(x) - c_1\right) \\ & + n \sum_{j\leqq n} \left(\int h(X_j, y)\,dF(y) - c_1\right). \end{aligned}$$

Here $c_1 = \int\int h(x,y)\,dF(x)\,dF(y)$ and $c_2 = \int h(x,x)\,dF(x)$. (1.12) follows immediately from (1.2) and (1.3). It is easy to see (Lemma 2.5 below) that if $\|h\|_2 < \infty$ then (as a matter of fact the argument used to prove Corollary 2 below shows that $\int\int h^2(x,y)\,dF(x)\,dF(y) < \infty$ is sufficient for the following discussion)

(1.13) $$V_n(h) \ll n \log n \quad \text{a.s.}$$

The law of the iterated logarithm yields

(1.14) $$\sum_{j\leqq n} (h(X_j, X_j) - c_2) \ll (n \log\log n)^{\frac{1}{2}} \quad \text{a.s.}$$

Consequently, if $\mathrm{Var}(\int h(x, X)dF(x)) \neq 0$ or if $\mathrm{Var}(\int h(X, y)dF(y)) \neq 0$, we can in (1.12) discard $V_n(h)$ and the left side of (1.14) since the last terms in (1.12) have order of magnitude $n^{\frac{3}{2}}$ and are thus the dominating terms. Yet as these dominating terms are sums of independent identically distributed random variables results on U-statistics in the case of nonvanishing variances fall into the domain of a well-developed theory.

This fact and several of the applications mentioned below have led to the following definition. A function $h: \mathbb{R}^2 \to \mathbb{R}$ with $\|h\|_2 < \infty$ is called degenerate for F if for all $x, y \in \mathbb{R}$

$$\int h(s, y)dF(s) = \int h(x, t)dF(t) = 0. \tag{1.15}$$

Thus by (1.12) if h is degenerate with respect to F and if (1.7) holds then Theorem 2 and Corollary 1 immediately imply

$$U_n(h) - W_n(h^*) = o(n \log\log n) \quad \text{a.s.} \tag{1.16}$$

and

$$\limsup_{n\to\infty} (2n \log\log n)^{-1} |U_n(h)| = C(h^*) \quad \text{a.s.} \tag{1.17}$$

respectively. Also if $\|h\|_2 < \infty$ then (1.6) implies

$$n^{-1} \max_{m \leq n} |U_m(h) + mc_2 - W_m(h^*)| \to 0 \quad \text{in prob.}$$

and likewise under the hypotheses of Theorem 4 we obtain

$$U_n(h) + nc_2 - W_n(h^*) \ll n^{1-\lambda} \quad \text{a.s.}$$

As a matter of fact it is easy to see by standard arguments that the following result holds. Let $\Delta = \{(x, y)\colon 0 \leq x \neq y \leq 1\}$ be the complement of the diagonal of $[0, 1]^2$.

Corollary 2. *If h is degenerate with respect to F and $\int\int h^2(x, y)dF(x)dF(y) < \infty$ then* (1.16) *and* (1.17) *hold with h^* replaced by $h^* 1_\Delta$. Moreover we have*

$$n^{-1} \max_{m \leq n} |U_m(h) - W_m(h^* 1_\Delta)| \to 0 \quad \textit{in prob.}$$

Also under the hypotheses of Theorem 4 *but with* (1.9) *weakened to* $\int\int |h(x, y)|^{2+\delta} dF(x)dF(y) < \infty$ *we have*

$$U_n(h) - W_n(h^* 1_\Delta) \ll n^{1-\lambda} \quad \text{a.s.}$$

In the non-degenerate case invariance principles for the U-statistic can be found in [13, 14, 20] etc. These generalize and refine Hoeffding's [11] classical theorem on the asymptotic normality of $n^{-\frac{3}{2}}(U_n(h) - n(n-1)c_1)$.

However, many interesting situations lead to the degenerate case. Several applications are mentioned in Sects. 1 and 2 of Hall [10]. We shall add a few more in Sect. 8. In the degenerate case distribution invariance principles have been proved by Neuhaus [18], Hall [10] and Denker et al. [4]. No almost sure nor probability invariance principles appear to have been published so far.

Section 2 contains moment inequalities for von Mises and U-statistics. The proof of Theorem 3 including the value of $C(h)$ are given in Sect. 3. In Sect. 4 we give the proof of Theorem 4, and in Sect. 5 the proofs of Theorems 1 and 2. Possible extensions of these results are discussed in Sect. 6. Finally, in Sect. 7 we give two sets of sufficient conditions on functions h to satisfy (1.10).

2. Preliminaries

Throughout this paper we shall assume that h is a degenerate kernel. This is no loss of generality since if we set

$$h_1(x, y) = h(x, y) - \int h(s, y) dF(s) - \int h(x, t) dF(t) + \int\int h(s, t) dF(s) dF(t)$$

then h_1 is a degenerate kernel. Also h_1 satisfies $\|h_1\|_{2+\delta} < \infty$, $\delta \geqq 0$ and (1.10) if h does. Moreover, $V_n(h_1) = V_n(h)$ and $W_t(h_1) = W_t(h)$.

We need to introduce some notation. If $A \subset \mathbb{R}$ is a measurable set and $L \subset \mathbb{N}$ is a finite set of integers we write

$$R(A, L) = \sum_{j \in L} (1\{X_j \in A\} - F(A)).$$

We also need the notion of a Kiefer process more general than the one introduced in Sect. 1. A separable Gaussian process $\{K(s,t), s \in \mathbb{R}, t \geqq 0\}$ is called a Kiefer process if $K(s, 0) = 0$, $s \in \mathbb{R}$, $\lim_{s \to \infty} K(s, t) = \lim_{s \to -\infty} K(s, t) = 0$, for all $t \geqq 0$, $EK(s, t) = 0$ for all $s \in \mathbb{R}$, $t \geqq 0$ and

$$EK(s, t) K(s', t') = (t \wedge t') F(s)(1 - F(s')) \qquad s \leqq s',\ t, t' \geqq 0$$

where F is a distribution function on $\mathbb{R}$. We note that if K is a standard Kiefer process and if K_F is a Kiefer process with respect to F, as just defined then with h^* as in Theorem 1

$$\int\int h(x, y) K_F(dx, m) K_F(dy, n) = \int\int h^*(x, y) K(dx, m) K(dy, n).$$

This follows from the definition. (See [4].) We also write

$$K(s, I) = K(s, n) - K(s, m), \qquad K(A, I) = \int 1_A(s) K(ds, I)$$

if $I = (m, n]$ for integers m, n and if $A \in \mathbb{R}$ is a measurable set.

For $\mathbf{a} = (a_1, a_2)$ and $\mathbf{n} = (n_1, n_2)$, $a_i, n_i \in \mathbb{Z}^+$, $i = 1, 2$ we set

$$S(\mathbf{a}, \mathbf{n}) = \sum h(X_i, X_j)$$

where the sum is extended over all i and j with $a_1 < i \leqq a_1 + n_1$, $a_2 < j \leqq a_2 + n_2$. Throughout this section we assume that $\|h\|_{2+\delta} < \infty$ for some $\delta \geqq 0$.

Lemma 2.1. $ES^2(\mathbf{a}, \mathbf{n}) \leqq 2 n_1 n_2 \|h\|_2^2$.

Proof. We have

$$ES^2(\mathbf{a}, \mathbf{n}) = \sum Eh(X_i, X_j) h(X_k, X_l)$$

where the sum is extended over all i, j, k, l with $a_1 < i, k \leqq a_1 + n_1$ and $a_2 < j, l \leqq a_2 + n_2$. By Fubini's theorem and (1.15) all terms in the sum vanish for which one index is different from the other three. The lemma follows now easily from these remarks.

Lemma 2.2. *Let $0 \leqq \delta \leqq 1$. Then there is a constant A, depending only on δ such that*

$$E|S(\mathbf{a}, \mathbf{n})|^{2+\delta} \leqq A \|h\|_{2+\delta}^{2+\delta} (n_1 n_2)^{1+\frac{1}{2}\delta}.$$

Proof. Without loss of generality we can assume that $\|h\|_{2+\delta} \leqq 1$. We shall first prove the inequality under the additional assumption

$$(2.1) \qquad a_1 + n_1 \leqq a_2 \quad \text{or } a_2 + n_2 \leqq a_1.$$

Recall that there exists a constant c_0, depending only on δ, such that: If $\{Y_j, j \geqq 1\}$ is a sequence of independent identically distributed random variables with $EY_1 = 0$ and $E|Y_1|^{2+\delta} < \infty$ then for all integers $m \geqq 0$, $n \geqq 1$

$$(2.2) \qquad E\left|\sum_{j=m+1}^{m+n} Y_j\right|^{2+\delta} \leqq c \; n^{1+\frac{1}{2}\delta} E|Y_1|^{2+\delta}.$$

Applying Fubini's theorem and (2.2) twice we obtain

$$(2.3) \qquad E|S(\mathbf{a}, \mathbf{n})|^{2+\delta} \leqq c^2 n_1^{1+\frac{1}{2}\delta} n_2^{1+\frac{1}{2}\delta} E|h(X_1, X_2)|^{2+\delta}.$$

We now prove the desired inequality for

$$a_1 = a_2 \quad \text{and} \quad n_1 = n_2 = n.$$

We follow ideas of Doob [6, p. 226f]. Without loss of generality we can assume $a_1 = a_2 = 0$. It is enough to prove that there exists a constant C such that

$$(2.4) \qquad E|S(\mathbf{0}, \mathbf{n})|^{2+\delta} \leqq C n^{2+\delta}$$

implies

$$(2.5) \qquad E|S(\mathbf{0}, 2\mathbf{n})|^{2+\delta} \leqq C(2n)^{2+\delta}, E|S(\mathbf{0}, 2\mathbf{n}+\mathbf{1})|^{2+\delta} \leqq C(2n+1)^{2+\delta}.$$

From this (2.4) will follow for all $\mathbf{n} = (n, n)$ by induction.

Now

$$\begin{aligned} S(\mathbf{0}, 2\mathbf{n}) &= S(\mathbf{0}, \mathbf{n}) + S((0, n); \mathbf{n}) + S(\mathbf{n}, \mathbf{n}) + S((n, 0); \mathbf{n}) \\ &= S_1 + S_2 + S_3 + S_4. \end{aligned}$$

Note that S_2 and S_4 satisfy (2.1) and that S_1 and S_3 are independent. Now

$$\begin{aligned} E|S_1 + S_3|^{2+\delta} &\leqq E\{(S_1 + S_3)^2 (|S_1|^\delta + |S_3|^\delta)\} \\ &\leqq E\{|S_1|^{2+\delta} + |S_3|^{2+\delta} + 2|S_1||S_3|^{1+\delta} \\ &\quad + 2|S_1|^{1+\delta}|S_3| + S_1^2 |S_3|^\delta + |S_1|^\delta S_3^2\} \\ &\leqq 2E|S_1|^{2+\delta} + 4E|S_1| \, E|S_1|^{1+\delta} + 2ES_1^2 E|S_1|^\delta. \end{aligned}$$

By Hölder's inequality and Lemma 2.1 we get for $\eta=\delta$, 1 and $1+\delta$

$$E|S_1|^\eta \leqq (2n^2)^{\frac{1}{2}\eta}.$$

Hence

$$E|S_1+S_3|^{2+\delta} \leqq 2E|S_1|^{2+\delta} + 12\cdot 2^{\frac{1}{2}\delta} n^{2+\delta}.$$

Consequently we obtain using Minkowski's inequality and (2.3)

$$\begin{aligned}(2.6)\qquad E|S(\mathbf{0},2\mathbf{n})|^{2+\delta} &\leqq ((E|S_1+S_3|^{2+\delta})^{1/(2+\delta)} + (E|S_2|^{2+\delta})^{1/(2+\delta)} \\ &\quad + (E|S_4|^{2+\delta})^{1/(2+\delta)})^{2+\delta} \\ &\leqq C(2n)^{2+\delta}\end{aligned}$$

if C is chosen so large that $2^{1/(2+\delta)} + (12\cdot 2^{\frac{1}{2}\delta}/C)^{1/(2+\delta)} + 5(c^2/C)^{1/(2+\delta)} \leqq 2$. This proves the first part of (2.5). To prove the second part we write

$$S(\mathbf{0},2\mathbf{n}+\mathbf{1}) = S(\mathbf{0},2\mathbf{n}) + S((2n,0),(1,2n)) + S((0,2n),(2n,1)) + S(2\mathbf{n},\mathbf{1}).$$

Hence by Minkowski's inequality, (2.3) and (2.6)

$$\begin{aligned}E|S(\mathbf{0},2\mathbf{n}+\mathbf{1})|^{2+\delta} &\leqq ((E|S(\mathbf{0},2\mathbf{n})|^{2+\delta})^{1/(2+\delta)} + 2c^{2/(2+\delta)}(2n)^{\frac{1}{2}}+1)^{2+\delta} \\ &\leqq C(2n)^{2+\delta} < C(2n+1)^{2+\delta}\end{aligned}$$

by our choice of C.

The case of general $\mathbf{a}$ and $\mathbf{n}$ follows now easily. The set of summation indices are lattice points in a rectangle with vertices (a_1,a_2), (a_1+n_1,a_2), (a_1+n_1,a_2+n_2), (a_1,a_2+n_2). We can decompose this rectangle into a square (possibly empty) whose diagonal lies on the 45 degree line and three rectangles (some possibly empty) all lying either above or below the 45 degree line. According to this decomposition we write

$$S(\mathbf{a},\mathbf{n}) = T_1+T_2+T_3+T_4, \qquad \text{say}$$

where $T_1=\sum h(X_i,X_j)$ with $a<i, j\leqq a+m$ for some a and $m\leqq n$ and T_2, T_3 and T_4 are such sums considered in (2.1). Hence by Minkowski's inequality

$$\begin{aligned}E|S(\mathbf{a},\mathbf{n})|^{2+\delta} &\leqq (C^{1/(2+\delta)} + 3c^{2/(2+\delta)})^{2+\delta}(n_1n_2)^{1+\delta/2} \\ &= A(n_1n_2)^{1+\delta/2}. \quad \square\end{aligned}$$

We also need a maximal inequality. For $\mathbf{a}=(a_1,a_2)$, $\mathbf{n}=(n_1,n_2)$ let

$$M(\mathbf{a},\mathbf{n}) = \max\{|S(\mathbf{a},\mathbf{p})|: 1\leqq p_1\leqq n_1, 1\leqq p_2\leqq n_2, p=(p_1,p_2)\}.$$

The following lemma follows immediately from Lemma 2.2 and Theorem 8 of Moricz [16].

Lemma 2.3. *We have for* $0\leqq\delta\leqq 1$

$$E(M(\mathbf{a},\mathbf{n}))^{2+\delta} \leqq A(n_1n_2\|h\|_{2+\delta}\log 2n_1 \log 2n_2)^{1+\frac{1}{2}\delta}.$$

For $\mathbf{a}=(a_1, a_2)$, $\mathbf{n}=(n_1, n_2)$ we write

$$(2.7)\qquad T(\mathbf{a},\mathbf{n})=\int\int h(x,y)K(dx,(a_1,a_1+n_1])K(dy,(a_2,a_2+n_2])$$

where K is the standard Kiefer process and

$$N(\mathbf{a},\mathbf{n})=\max\{|T(\mathbf{a},\mathbf{p})|: 1\leqq p_1\leqq n_1, 1\leqq p_2\leqq n_2\}.$$

Lemma 2.4. *Lemma* 2.3 *holds with M replaced by N.*

Proof. By (1.4) and (2.7) we have

$$T(\mathbf{a},\mathbf{n})=\sum\int\int h(x,y)B_i(dx)B_j(dy)=\sum Z_{ij}, \qquad \text{say}$$

where similar to the definition of $S(\mathbf{a},\mathbf{n})$ the sums are extended over all i and j with $a_1<i\leqq a_1+n_1$, $a_2<j\leqq a_2+n_2$. Hence and by [16, Theorem 8] for the proof of the lemma it suffices to show that

$$(2.8)\qquad E|T(\mathbf{a},\mathbf{n})|^{2+\delta}\leqq A(n_1 n_2)^{1+\frac{1}{2}\delta}.$$

This can be proved in the same way with the argument given in the proof of Lemma 2.2. Since $\mathscr{L}(n^{-1}T(\mathbf{0},(n,n)))=\mathscr{L}(T(\mathbf{0},\mathbf{1}))$ by (2.7) and since

$$(2.9)\qquad \mathscr{L}(r^{-1}V_r(h))\to\mathscr{L}(T(\mathbf{0},\mathbf{1}))$$

by [8] we obtain from Fatou's lemma and Lemma 2.2

$$(2.10)\qquad \begin{aligned} n^{-2-\delta}E|T(\mathbf{0},\mathbf{n})|^{2+\delta}&=E|T(\mathbf{0},\mathbf{1})|^{2+\delta}\\ &\leqq\liminf_{r\to\infty} r^{-2-\delta}E|V_r(h)|^{2+\delta}\\ &\leqq A\,\|h\|_{2+\delta}^{2+\delta}.\end{aligned}$$

This shows that $E|Z_{11}|^{2+\delta}<\infty$. In the same way one can show that $E|Z_{12}|^{2+\delta}<\infty$. For this we replace (2.9) by

$$\mathscr{L}(r^{-1}\int\int h(x,y)R(dx,(n,2n])R(dy,(0,n])\to\mathscr{L}(Z_{12})$$

which follows from [4, Theorem 5] applied with $K=2$, $\alpha_1=\alpha_2=\frac{1}{2}$, $m_1=m_2=1$ and $t_1=t_2=1$. Hence we obtain as in the proof of (2.3) that (2.8) holds if (2.1) is satisfied. (2.10) replaces (2.4). The proof of (2.8) in the general case finally can be completed in the same way as the proof of Lemma 2.2. □

Lemma 2.5. *Let h: $\mathbb{R}^2\to\mathbb{R}$ with $\|h\|_2<\infty$, degenerate or non-degenerate. Then as $n\to\infty$*

$$V_n(h)\ll n\log n \qquad \text{a.s.}$$

Remark. Lemma 2.3 immediately yields the bound $n\log^3 n$.

Proof. Recall that we can assume without loss of generality that h is degenerate. Hence by (1.12)

$$(2.11)\qquad V_n(h)=U_n(h)+\sum_{j\leqq n} h(X_j,X_j).$$

Let $\mathscr{L}_n$ be the σ-field generated by $X_1, \ldots, X_n$. Then $\{V_n(h) - n\int h(x,x)dF(x), \mathscr{L}_n, n \geqq 1\}$ is a martingale by [14, Lemma 2.1]. (Here is the quick proof: Since $E\{h(X_n, X_n)|\mathscr{L}_{n-1}\} = Eh(X_1, X_1) = \int h(x,x)dF(x)$ and since for $i<n$

$$E\{h(X_i, X_n)|\mathscr{L}_{n-1}\} = \int h(X_i, y)dF(y) = 0$$

by independence the claim follows from (1.12).) Hence by Doob's inequality (the martingale version of Kolmogorov's inequality) and by Lemma 2.1 we have

$$\begin{aligned} P\{\max_{m \leqq 2^k} |V_m(h) - m\int h(x,x)dF(x)| \geqq 2^k k\} \\ \ll 2^{-2k}k^{-2}(E(V_{2^k}(h))^2 + 2^{2k}\|h\|_2^2) \ll k^{-2}. \end{aligned}$$

Thus by the Borel Cantelli lemma $\max_{m \leqq 2^k} |V_m(h)| \ll 2^k k$ a.s. Now let n be given. Find k such that $2^{k-1} < n \leqq 2^k$. Then

$$|V_n(h)| \leqq \max_{m \leqq 2^k} |V_m(h)| \ll 2^k k \ll n \log n \quad \text{a.s.}$$

3. Proof of Theorem 2

We first prove an exponential bound for $W_1(h)$.

Theorem 5. *Let K and h be as in Theorem 3. Then for all* $0 < t < (4\|h\|_2)^{-1}$

$$E\exp(tW_1(h)) \leqq \exp(2t^2\|h\|_2^2 + t\|h\|_2). \tag{3.1}$$

Proof. We first prove (3.1) under the additional assumptions that h is symmetric, i.e. $h(x,y) = h(y,x)$, $0 \leqq x, y \leqq 1$ and that h vanishes on the diagonal, i.e. $h(x,x) = 0$, $0 \leqq x \leqq 1$. Then it follows from [8] and [18] that $W_1(h)$ can be represented in the form

$$W_1(h) \stackrel{\mathscr{L}}{=} \sum_{i \geqq 1} \lambda_i(N_i^2 - 1)$$

where $\{N_i, i \geqq 1\}$ is a sequence of independent standard normal random variables and the λ_i's are constants satisfying $\sum_{i \geqq 1} \lambda_i^2 \leqq \|h\|_2^2$. Thus

$$\begin{aligned} E\exp(tW_1(h)) &= \prod_{i \geqq 1} E\exp(t\lambda_i(N_i^2 - 1)) \\ &= \prod_{i \geqq 1} \exp(-t\lambda_i)(1 - 2t\lambda_i)^{-\frac{1}{2}} \\ &= \prod_{i \geqq 1} \exp(-t\lambda_i - \tfrac{1}{2}\log(1 - 2t\lambda_i)) \\ &\leqq \exp(\tfrac{1}{2}\sum_{i \geqq 1} 4t^2\lambda_i^2) \leqq \exp(2t^2\|h\|_2^2) \end{aligned}$$

since $|x + \log(1-x)| \leqq x^2$ for $|x| \leqq \frac{1}{2}$. This proves (3.1) under the additional assumptions that h is symmetric and vanishes on the diagonal.

To remove these extra assumptions we put $\Delta=\{(x, y): 0 \leqq x \neq y \leqq 1\}$ and

(3.2) $$f=h 1_{\Delta}, \quad g=h-f.$$

Then

(3.3) $$W_1(h)=W_1(f)+W_1(g).$$

We now prove that

(3.4) $$W_1(g)=\int h(x, x) d x \quad \text{a.s.}$$

To see this we note that g is degenerate with respect to G, the uniform distribution on $[0,1]$. Hence by (2.9)

$$\mathscr{L}(r^{-1} V_r(g)) \rightarrow \mathscr{L}(W_1(g)).$$

But by the strong law of large numbers we have with probability 1

$$r^{-1} V_r(g)=r^{-1} \sum_{i \leqq r} g(u_i, u_i) \rightarrow \int h(x, x) d x.$$

Here $\{u_i, i \geqq 1\}$ is a sequence of independent random variables uniformly distributed over $[0,1]$. These two relations imply (3.4). Next define h' by

$$h'(x, y)=\tfrac{1}{2}(h(x, y) 1_{\Delta}(x, y)+h(y, x) 1_{\Delta}(y, x)).$$

Then h' is symmetric and vanishes on the diagonal. Moreover, by an easy calculation $W_1(h')=W_1(f)$ and $\|h'\|_2 \leqq \|h\|_2$. Hence by (3.3), (3.4) and the special case already proved

$$\begin{aligned} E \exp(t W_1(h)) & =\exp(t W_1(g)) E \exp(t W_1(f)) \\ & \leqq \exp(t\|h\|_2+2 t^2\|h\|_2^2). \quad \square \end{aligned}$$

Next we prove a crude version of Corollary 1.

Lemma 3.1. *Let K and h be as in Theorem 3. Then with probability 1*

$$\limsup_{n \rightarrow \infty}(n \log \log n)^{-1} W_n(h) \leqq 20\|h\|_2.$$

Proof. Recall that by (1.4)

$$W_n(h)=\sum_{i, j \leqq n} \iint h(x, y) B_i(d x) B_j(d y).$$

Let $\mathscr{F}_n$ be the σ-field generated by $B_1, \ldots, B_n$ and set

$$c=E \iint h(x, y) B_1(d x) B_1(d y).$$

Then $\{W_n(h)-c n, \mathscr{F}_n, n \geqq 1\}$ is a martingale. Hence by Doob's maximal inequality for submartingales and by Theorem 5 with $t=(5\|h\|_2)^{-1}$ we have for each $k \geqq 1$

$$
\begin{aligned}
&P\{\max_{n\leqq 2^k}(W_n(h)-nc)\geqq 10\cdot 2^k\|h\|_2\log\log 2^k\}\\
&\quad=P\{\max_{n\leqq 2^k}\exp(2^{-k}t(W_n(h)-nc))\geqq\exp(2\log\log 2^k)\}\\
&\quad\leqq\exp(-2\log\log 2^k)E\exp(2^{-k}t(W_{2^k}(h)-2^kc))\\
&\quad\ll k^{-2}E\exp(tW_1(h))\ll k^{-2}
\end{aligned}
$$

since $\mathscr{L}(n^{-1}W_n(h))=\mathscr{L}(W_1(h))$ for all $n\geqq 1$. The lemma follows now from the Borel Cantelli lemma. □

We now start with the proof of Theorem 3. Recall that G denotes the distribution function of the uniform distribution on $[0,1]$. Let λ be the Lebesgue measure on $[0,1]$ and let $\phi:[0,1]^2\to\mathbb{R}$ be a measurable function in $L^2(\lambda\times\lambda)$. Define an operator $A_\phi: L^2(\lambda)\to L^2(\lambda)$, associated with ϕ, by setting

$$A_\phi(f)(x)=\int_0^1 f(y)\phi(x,y)\,dy.$$

Then A_ϕ is a Hilbert-Schmidt operator and if ϕ is symmetric, i.e. $\phi(x,y)=\phi(y,x)$ then A_ϕ is self-adjoint and we have

$$\|A_\phi\|=\sup_{\|f\|=1}\langle A_\phi f,f\rangle=\max\{|\mu|:\mu \text{ eigenvalue of } A_\phi\}. \tag{3.5}$$

Next, let $h:[0,1]^2\to\mathbb{R}$ be such that $\|h\|_2<\infty$ where $\|\ \|_2$ is defined in (1.1) but with F replaced by G. Let h_1 be as in the beginning of Sect. 2, but also with F replaced by G, i.e.

$$h_1(x,y)=h(x,y)-\int_0^1 h(s,y)\,ds-\int_0^1 h(x,t)\,dt+\int_0^1\int_0^1 h(s,t)\,ds\,dt.$$

Finally, let

$$\hat{h}(x,y)=\tfrac{1}{2}(h_1(x,y)+h_1(y,x)).$$

Then $\hat{h}$ is symmetric and degenerate for G.
The following lemma identities the limit in Theorem 3.

Lemma 3.2. *Let h be as above with $\|h\|_2<\infty$. Then with probability* 1

$$\limsup_{n\to\infty}(2n\log\log n)^{-1}|W_n(h)|=\|A_{\hat{h}}\|.$$

Proof. We first note that $W_t(h)=W_t(\hat{h})$. Moreover, by (3.4) and since

$$\mathscr{L}((t^{-\frac{1}{2}}K(s,t))_{0\leqq s\leqq 1})=\mathscr{L}((K(s,1))_{0\leqq s\leqq 1})\qquad\text{for all } t>0$$

$$W_t(g)\overset{\mathscr{L}}{=}t^{-1}W_1(g)=t\int_0^1 h(x,x)\,dx. \tag{3.6}$$

Hence by (3.3) and since $A_g\equiv 0$ we can assume for the proof of the lemma without loss of generality that h is symmetric, vanishes on the diagonal, is degenerate for G and satisfies $\|h\|_2<\infty$.

Now h can be represented in $L^2(\lambda\times\lambda)$ in the form

$$h(x,y)=\sum_{j\geqq 1}\mu_j f_j(x)f_j(y) \tag{3.7}$$

where μ_j are the eigenvalues of the operator A_h and where $\{f_j, j\geqq 1\}$ is a system of corresponding eigenfunctions orthonormal with respect to $L^2(\lambda)$. Since h vanishes on the diagonal we have $\sum \mu_j^2 = \|h\|_2^2$ and the processes $\{W_n(h), n\geqq 0\}$ and $\{\sum_{j\geqq 1} \mu_j(Y_j^2(n)-n), n\geqq 0\}$ have the same laws. (See [4], Lemma 7.) Here $\{Y_j, j\geqq 1\}$ is a sequence of independent standard Brownian motions. Also by (3.5)

$$\|A_h\| = \max_{j\geqq 1} |\mu_j|. \tag{3.8}$$

We now set

$$\begin{aligned} h_k(x,y) &= \sum_{j\leqq k} \mu_j f_j(x) f_j(y) \qquad &\text{if } x \neq y \\ &= 0 &\text{if } x = y. \end{aligned}$$

Let $\varepsilon > 0$. Then by (3.7) $\|h-h_k\|_2 < \varepsilon/20$ and

$$\|A_h - A_{h_k}\| < \varepsilon \tag{3.9}$$

if k is sufficiently large. Consequently and by Lemma 3.1 we get

$$\limsup_{n\to\infty} (2n \log\log n)^{-1} |W_n(h-h_k)| < \varepsilon \quad \text{a.s.} \tag{3.10}$$

Now since $\{W(h_k), n\geqq 0\}$ and $\{\sum_{j\leqq k} \mu_j(Y_j^2(n)-n), n\geqq 0\}$ have the same laws we have with probability 1

$$\begin{aligned} &\limsup_{n\to\infty} (2n\log\log n)^{-1} |W_n(h_k)| \\ &\quad = \limsup_{n\to\infty} (2n\log\log n)^{-1} \Big|\sum_{j\leqq k} \mu_j (Y_j^2(n)-n)\Big| \\ &\quad = \limsup_{n\to\infty} (2n\log\log n)^{-1} \Big|\sum_{j\leqq k} \mu_j Y_j^2(n)\Big| \\ &\quad = \sup\Big\{\Big|\sum_{j\leqq k} \mu_j x_j^2\Big| : \sum_{j\leqq k} x_j^2 \leqq 1\Big\} = \max_{j\leqq k} |\mu_j| = \|A_{h_k}\| \end{aligned}$$

by the compact law of the iterated logarithm for standard $\mathbb{R}^k$-valued Brownian motion. The lemma follows now from (3.9) and (3.10). □

4. Proof of Theorem 4

We need the following trivial fact.

Lemma 4.1. *We can assume without loss of generality that the* $h(i,j,r)$ *in* (1.10) *satisfy*

$$|h(i,j,r)| \leqq 2^{r\gamma/\delta}.$$

Proof. This follows immediately from (1.10) since

$$\iint |h(x,y)|^2 1\{|h(x,y)| > d^{\gamma/\delta}\} dF(x)dF(y) \leqq d^{-\gamma} \iint |h(x,y)|^{2+\delta} dF(x)dF(y)$$

and

$$\int |h(x,x)|^2 1\{|h(x,x)|>d^{\gamma/\delta}\}\, dF(x) \leqq d^{-\gamma} \int |h(x,x)|^{2+\delta} dF(x)$$

upon setting $d=2^r$. □

Recall that by the remark at the beginning of Sect. 2 we can assume without loss of generality that the kernel h is degenerate for F. Thus

$$\int\int h(x,y) R(dx, L) R(dy, M) = \sum_{i\in L,\, j\in M} h(X_i, X_j).$$

We also recall the definition of a general Kiefer process K_F. Throughout this section, however, we will suppress the index F in K_F unless stated otherwise.

For convenience we introduce some notation. In addition to α and λ, defined in (1.11) we set

$$\eta = 4/\gamma \tag{4.1}$$

and

$$t_k = t(k) = [k^\alpha], \qquad H_k = (t_{k-1}, t_k] \cap \mathbb{Z}, \qquad n_k = \operatorname{card} H_k, \qquad k=1,2,\dots. \tag{4.2}$$

Thus

$$k^{\alpha-1} \ll n_k \ll k^{\alpha-1}. \tag{4.3}$$

Moreover, we write

$$r_k = [\eta \log k / \log 2], \qquad d_k = 2^{r_k} \tag{4.4}$$

so that

$$d_k \leqq k^\eta. \tag{4.5}$$

Before presenting the details we shall give an outline of the proof of Theorem 4. On H_k we partition the real line into d_k sets $A(i,k) = A_{ir_k}$, $1 \leqq i \leqq d_k$ where A_{ir} are chosen according to (1.10). Next, we define the "skeleton process" $\{R_k, k \geqq 1\}$ of the empirical process R by

$$R_k(i) = R(A(i,k), H_k) \qquad 1 \leqq i \leqq d_k. \tag{4.6}$$

R_k is a sum of n_k independent identically distributed random vectors with values in $\mathbb{R}^{d_k}$ and hence by the multivariate central limit theorem close to a normal distribution. We then can apply a result of [19] to obtain an almost sure approximation of R_k by Gaussian random vectors $Y_k = \{Y_k(i), 1 \leqq i \leqq d_k\}$. By a simple measure theoretic argument we then can choose a Kiefer process K such that $Y_k = K(A(i,k), H_k)$, $1 \leqq i \leqq d_k$, $k=1,2,\dots$. This process has the desired properties. In order to prove this we shall first use Lemma 2.3 to show that

$$\max_{t_k < m \leqq t_{k+1},\, t_l < n \leqq t_{l+1}} \left\{\int\int h(x,y) R(dx,m) R(dy,n) - \int\int h(x,y) R(dx,t_k) R(dy,t_l)\right\}$$

and the same expression with R replaced by K are sufficiently small. This reduces the problem to estimating the difference

$$\iint h(x,y)R(dx,t_k)R(dy,t_k)-\iint h(x,y)K(dx,t_k)K(dy,t_k).$$

In the next step we reduce this once more using Lemma 2.2 to the estimation of (here $\kappa=\kappa(k)=[k^{\frac{1}{2}}]$)

$$\iint h(x,y)R(dx,(t_\kappa,t_k])R(dy,(t_\kappa,t_k])-\iint h(x,y)K(dx,(t_\kappa,t_k])K(dy,(t_\kappa,t_k]).$$

In these integrals we can replace h by a suitable step function and using (1.10) we subsequently can control the error introduced. The stochastic integrals over these step functions can be represented as sums involving R and K and thus their difference can be estimated without much difficulties.

We shall now present the details of the proof. By (4.6) and the definition of R we have

$$R_k(i)=\sum_{n\in H_k}(1\{X_n\in A(i,k)\}-F(A(i,k))),\qquad 1\leqq i\leqq d_k. \tag{4.7}$$

Hence R_k is a sum of independent identically distributed random vectors with mean 0 and covariance matrix $C_k=((c_{ij}(k)))$ where

$$\begin{aligned} c_{ij}(k)&=-F(A(i,k))F(A(j,k)) && \text{if } i\neq j\\ &=F(A(i,k))(1-F(A(i,k))) && \text{if } i=j. \end{aligned} \tag{4.8}$$

We apply Yurinskii's theorem [22] and get for the Prohorov distance

$$\pi(\mathscr{L}(n_k^{-\frac{1}{2}}R_k),\mathscr{N}(O,C_k))\ll n_k^{-\frac{1}{9}}d_k^{\frac{1}{3}}\ll k^{-(\alpha-1)/9+\eta/3}. \tag{4.9}$$

Here $\mathscr{N}(O,C_k)$ denotes the Gaussian law with mean zero and covariance matrix C_k. Hence in view of (4.9) we obtain applying [19, Theorem 3] without loss of generality a sequence $\{Y_k,k\geqq 1\}$ of independent $\mathscr{N}(O,C_k)$-distributed random vectors such that

$$P\{\|n_k^{-\frac{1}{2}}R_k-Y_k\|\geqq Ck^{-(\alpha-1)/9+\eta/3}\}\ll k^{-(\alpha-1)/9+\eta/3}\ll k^{-2}$$

using (4.1). Here C is a positive constant implied by $\ll$ in (4.9). The Borel Cantelli lemma yields as $k\to\infty$

$$\|n_k^{-\frac{1}{2}}R_k-Y_k\|\ll k^{-(\alpha-1)/9+\eta/3}\qquad \text{a.s.} \tag{4.10}$$

As is easily seen the sequences $\{n_k^{-\frac{1}{2}}K(A(i,k),H_k),\ 1\leqq i\leqq d_k,\ k\geqq 1\}$ and $\{Y_k,k\geqq 1\}$ have the same law. Hence by [2, Lemma A1] we can assume

$$Y_k(i)=n_k^{-\frac{1}{2}}K(A(i,k),H_k),\qquad 1\leqq i\leqq d_k,\qquad k\geqq 1 \tag{4.11}$$

where $Y_k(i)$ denotes the i-th component of Y_k. Hence by (4.10) and (4.11) we get with probability 1

$$\begin{aligned}\Big(\sum_{i\leqq d_k}(R(A(i,k),H_k)-K(A(i,k),H_k))^2\Big)^{\frac{1}{2}}&\ll k^{-(\alpha-1)/9+\eta/3}n_k^{\frac{1}{2}}\\ &\ll k^{(7\alpha+2)/18+\eta/3}.\end{aligned} \tag{4.12}$$

Let $\kappa=[k^{\frac{1}{2}}]$

(4.13) $$I_k=(t_\kappa, t_k]$$

and

(4.14) $$g_k(x,y)=\sum_{1\leqq i,j\leqq d_\kappa} h(i,j,r_\kappa)1\{x\in A(i,\kappa)\}1\{y\in A(j,\kappa)\}.$$

Lemma 4.2. *We have with probability 1*

(4.15) $$\iint g_k(x,y)(R(dx,I_k)R(dy,I_k)-K(dx,I_k)K(dy,I_k))\ll t_k^{17/18}.$$

Proof. Since $I_k=\bigcup_{l=\kappa+1}^{k} H_l$ we have using (4.14)

$$\iint g_k(x,y)R(dx,I_k)R(dy,I_k)=\sum_{\kappa<l,m\leqq k}\iint g_k(x,y)R(dx,H_l)R(dy,H_m)$$
$$=\sum_{\kappa<l,m\leqq k}\sum_{1\leqq i,j\leqq d_\kappa} h(i,j,r_\kappa)R(A(i,\kappa),H_l)R(A(j,\kappa),H_m).$$

Writing the stochastic integral with respect to the Kiefer process in the same way we can rewrite the left side of (4.15) in the form

(4.16) $$\sum_{\kappa<l,m\leqq k}\Big\{\sum_{1\leqq i,j\leqq d_\kappa} h(i,j)(R(A(i),H_l)-K(A(i),H_l))R(A(j),H_m)$$
$$+\sum_{1\leqq i,j\leqq d_\kappa} h(i,j)(R(A(j),H_m)-K(A(j),H_m))K(A(i),H_l)\Big\}.$$

Here we dropped κ in h and A. By (4.5), (4.13) and Lemma 4.1

(4.17) $$\max_{1\leqq i,j\leqq d}|h(i,j,r_\kappa)|\ll k^{\frac{1}{2}\eta\gamma/\delta}\ll k^{2/\delta}.$$

Thus the last inner sum in (4.16) is

(4.18) $$\ll k2/\delta\sum_{j\leqq d_\kappa}|R(A(j),H_m)-K(A(j),H_m)|\sum_{i\leqq d_\kappa}|K(A(i),H_l)|.$$

By (4.1), Cauchy's inequality, (4.12) and (4.5)

(4.19) $$\sum_{j\leqq d_\kappa}|R(A(j,\kappa),H_m)-K(A(j,\kappa),H_m)|$$
$$\leqq\sum_{j\leqq d_m}|R(A(j,m),H_m)-K(A(j,m),H_m)|$$
$$\leqq d_m^{\frac{1}{2}}\Big(\sum_{j\leqq d_m}(R(A(j,m),H_m)-K(A(j,m),H_m))^2\Big)^{\frac{1}{2}}$$
$$\ll m^{\frac{1}{2}\eta}m^{(7\alpha+2)/18+(\eta/3)}\ll k^{(7\alpha+2)/18+5\eta/6}\quad\text{a.s.}$$

To estimate the last sum in (4.18) we define the random vector $K_l=(K(A(i,l),H_l),\ 1\leqq i\leqq d_l)$. Since $n_l^{-\frac{1}{2}}K_l$ is Gaussian with mean zero and covariance matrix C_l as defined in (4.8), we obtain $E\|n_l^{-\frac{1}{2}}K_l\|^2=\operatorname{tr} C_l<1$ since

$\{A(i,l),\ 1 \leqq i \leqq d_l\}$ is a partition of the real line. Hence by the Fernique-Landau-Shepp inequality [7] there is a constant $c>0$ such that

$$P\{n_l^{-\frac{1}{2}} \|K_l\| > \rho\} \leqq \exp(-c\rho^2), \qquad \rho \geqq 1.$$

We set $\rho = l^{\frac{1}{2}}$ and use the Borel Cantelli lemma and (4.3) to get

$$\|K_l\| \ll n_l^{\frac{1}{2}} l^{\frac{1}{2}} \ll l^{\frac{1}{2}\alpha} \qquad \text{a.s.} \tag{4.20}$$

Since

$$\sum_{i \leqq d_\kappa} |K(A(i), H_l)| \leqq \sum_{i \leqq d_l} |K(A(i,l), H_l)| \leqq d_l^{\frac{1}{2}} \|K_l\|$$

we obtain from (4.17)-(4.20) that with probability 1

$$\begin{aligned}(4.21)\qquad & \Big|\sum_{1 \leqq i,j \leqq d_\kappa} h(i,j)(R(A(j), H_m) - K(A(j), H_m))\, K(A(i), H_l)\Big| \\ & \ll d_k^{\frac{1}{2}} k^{2/\delta} k^{(7\alpha+2)/18 + 5\eta/6} k^{\frac{1}{2}\alpha} \ll k^{2/\delta + (8\alpha+1)/9 + 4\eta/3}.\end{aligned}$$

Since $|R(A(j), H_m)| \leqq |R(A(j), H_m) - K(A(j), H_m)| + |K(A(j), H_m)|$ we obtain by (4.17)-(4.20) the same estimate for the first inner sum in (4.16). Hence by (4.16), (4.21), (4.1), (4.2) and (1.11) we obtain the result.

Lemma 4.3. *We have with probability* 1

$$\max_{t_k < n \leqq t_{k+1}} |V_n(h) - V_{t_k}(h)| \ll t_k^{1-\lambda}.$$

Proof. The left side is bounded by $M((t_k, 0), (n_k, t_{k+1})) + M((0, t_k), (t_{k+1}, n_k))$. By symmetry it is enough to estimate just one of these quantities. Recall that we assume without loss of generality $\|h\|_{2+\delta} \leqq 1$. By Markov's inequality, Lemma 2.3, (4.2) and (4.3) we get

$$\begin{aligned} P\{M((t_k, 0), (n_k, t_{k+1})) \geqq t_k^{1-\lambda}\} & \ll t_k^{-(1-\lambda)(2+\delta)} (\log k)^{4+2\delta} (n_k t_k)^{1+\frac{1}{2}\delta} \\ & \ll k^{-\alpha(1-\lambda)(2+\delta)} k^{(2\alpha-1)(1+\frac{1}{2}\delta)} (\log k)^{4+2\delta} \\ & \ll k^{(\alpha\lambda - \frac{1}{2})(2+\delta)} (\log k)^{4+2\delta}. \end{aligned}$$

By (4.1) and (1.11) the exponent of k is less than -1. Hence we can apply the Borel Cantelli lemma and obtain the result.

Lemma 4.4. *We have with probability* 1

$$\iint h(x,y) R(dx, t_k) R(dy, t_\kappa) \ll t_k^{7/8} \tag{4.22}$$

and

$$\iint h(x,y) R(dx, t_\kappa) R(dy, t_\kappa) \ll t_k^{7/8}. \tag{4.23}$$

Proof. The left side of (4.22) equals $S(\mathbf{0}, (t_k, t_\kappa))$. By Markov's inequality, Lemma 2.2, (4.2) and (1.11) we have

$$P\{S(\mathbf{0}, (t_k, t_\kappa)) \geqq t_k^{7/8}\} \ll t_k^{-7/4} t_k t_\kappa \ll k^{-\frac{1}{4}\alpha} \ll k^{-5}.$$

The Borel Cantelli lemma immediately yields (4.22). Similarly the left side of (4.23) equals $S(\mathbf{0},(t_\kappa, t_\kappa))$ and the desired estimate follows in the same way.

Lemma 4.5. *We have with probability* 1

$$\int\int (h(x,y)-g_k(x,y))R(dx,(t_\kappa,t_k])R(dy,(t_\kappa,t_k])\ll t_k^{1-\lambda}.$$

Proof. By (4.14), (4.5) and (1.10) we have $\|h-g_k\|_2 \ll k^{-\frac{1}{4}\eta\gamma} \ll k^{-2}$. Define $S_k(\mathbf{a},\mathbf{n})$ in the same way as $S(\mathbf{a},\mathbf{n})$ but with h replaced by $h-g_k$. Then

$$\int\int (h(x,y)-g_k(x,y))R(dx,(t_\kappa,t_k])R(dy,(t_\kappa,t_k])=S_k(\mathbf{t}_\kappa,\mathbf{t}_k-\mathbf{t}_\kappa)=S_k.$$

By Chebyshev's inequality, Lemma 2.1, (4.1) and (4.3)

$$P\{|S_k|\geqq t_k^{1-\lambda}\}\ll t_k^{-2(1-\lambda)}t_k^2 k^{-2}\ll k^{-2+2\alpha\lambda}\ll k^{-\frac{3}{2}}.$$

The Borel Cantelli lemma yields the result.

Because of Lemma 2.4 we see that Lemmas 4.3–4.5 remain valid with R replaced by K. In view of the outline of the proof of Theorem 1 given at the beginning of this section we obtain Theorem 3 from Lemmas 4.2–4.5, from the adaptions of Lemmas 4.3–4.5 to the Kiefer process via seven applications of the triangle inequality and the fact that

$$\int\int h(x,y)K_F(dx,n)K_F(dy,n)=\int\int h^*(x,y)K_G(dx,n)K_G(dy,n) \tag{4.24}$$

as was observed at the beginning of Sect. 2.

5. Proof of Theorems 1 and 2

Before we start with the proofs we want to make several remarks. Proofs based on the representation of h in the form (3.7) presumably will not be any simpler than the ones given below, particularly, since we will use much of the material developed in Sect. 4. Moreover, proofs based on (3.7) do not lend themselves to a generalization of these theorems to kernels h in more than two variables.

We first prove a crude version of Corollary 1.

Proposition 5.1. *Under the hypotheses of Theorem* 1 *we have with probability* 1

$$\limsup_{n\to\infty}(n\log\log n)^{-1}V_n(h)\leqq 800\|h\|_2.$$

The proof of Proposition 5.1 will be given in a series of lemmas. We put

$$v_m=V_m-V_{m-1},\qquad m\geqq 1 \tag{5.1}$$

and denote by $\mathscr{L}_m$ the σ-field generated by $X_1,\ldots,X_m$.

Lemma 5.2. *We have with probability* 1

$$\limsup_{m\to\infty}(m\log\log m)^{-1}E(v_m^2|\mathscr{L}_{m-1})\leqq 6\|h\|_2^2. \tag{5.2}$$

Proof. Since h is degenerate we have by (2.11)

$$\tfrac{1}{3}v_m^2 \leqq \Big(\sum_{i<m} h(X_i, X_m)\Big)^2 + \Big(\sum_{j<m} h(X_m, X_j)\Big)^2 + h^2(X_m, X_m). \tag{5.3}$$

As $E\{h^2(X_m, X_m) | \mathscr{L}_{m-1}\} \leqq \|h\|_2^2$ we need to concentrate only on the first term in (5.3). By independence we have

$$\begin{aligned} E\Big\{\Big(\sum_{i<m} h(X_i, X_m)\Big)^2 \Big| \mathscr{L}_{m-1}\Big\} &= \int_0^1 \Big(\sum_{i<m} h(X_i, u)\Big)^2 du \\ &= \Big\|\sum_{i<m} h(X_i, \cdot)\Big\|_{L^2(F)}^2 \end{aligned}$$

The lemma follows now from the law of the iterated logarithm for sequences of random variables with values in the Hilbert space $L^2(F)$. (See e.g. [12, Theorem 4.1].) □

Next, we put

$$y_m = v_m 1\{|v_m| \leqq 50\|h\|_2 m\}, \qquad w_m = y_m - E(y_m | \mathscr{L}_{m-1}), \qquad m \geqq 1. \tag{5.4}$$

Lemma 5.3. *We have with probability* 1

$$\limsup_{n\to\infty} (n \log\log n)^{-1} \sum_{m \leqq n} w_m \leqq 600\|h\|_2.$$

Proof. For fixed $k \geqq 1$ the sequence $\{w_m, \mathscr{L}_m, 1 \leqq m \leqq 2^k\}$ is a martingale difference sequence uniformly bounded by $c = 100\|h\|_2 2^k$. We apply Lemma 5.4.1 and Corollary 5.4.1 of Stout [21] with $\lambda = 1/c$ and obtain

$$P\Big\{\max_{n \leqq 2^k} \exp\Big(\lambda \sum_{m \leqq n} w_m - \tfrac{1}{2}\lambda^2 \tfrac{3}{2} \sum_{m \leqq n} E(w_m^2 | \mathscr{L}_{m-1})\Big) > \tfrac{1}{4}k^2\Big\} \leqq 4k^{-2}.$$

Hence by the Borel Cantelli lemma there is with probability 1 a $k_0 = k_0(\omega)$ such that for all $k \geqq k_0$

$$\begin{aligned} \max_{n \leqq 2^k} \lambda \sum_{m \leqq n} w_m &\leqq 2\log\log 2^k + \tfrac{3}{4}\lambda^2 \sum_{m \leqq 2^k} E(w_m^2 | \mathscr{L}_{m-1}) \\ &\leqq 2\log\log 2^k + \tfrac{3}{4}\lambda^2 \sum_{m \leqq 2^k} E(v_m^2 | \mathscr{L}_{m-1}). \end{aligned} \tag{5.5}$$

Now Lemma 5.2 implies that there exists with probability 1 an $m_0 = m_0(\omega)$ such that for all $m \geqq m_0(\omega)$ and all ω

$$E(v_m^2 | \mathscr{L}_{m-1}) \leqq 12\|h\|_2^2 m \log\log m.$$

Hence by (5.5)

$$\max_{n \leqq 2^k} \lambda \sum_{m \leqq n} w_m \leqq 3\log\log 2^k.$$

We substitute λ and obtain the lemma. □

For the proof of Proposition 5.1 it remains to show that in Lemma 5.3 w_m can be replaced by v_m. This will follow from the following two lemmas and the strong law of large numbers applied to the sequence $\{h(X_n, X_n), n \geqq 1\}$.

Lemma 5.4. *With probability* 1 *there exists an* $m_0 = m_0(\omega)$ *such that for all* $m \geqq m_0$

$$|w_m - y_m| \leqq 2 \|h\|_2 \log\log m.$$

Proof. Recall from the proof of Lemma 2.4 that $\{v_m - \int h(x,x)\,dx, \mathscr{L}_m, m \geqq 1\}$ is a martingale difference sequence. Thus by (5.4) and Lemma 5.2 there exists with probability 1 an $m_0 = m_0(\omega)$ such that for all $m \geqq m_0$

$$\begin{aligned} |w_m - y_m| &= |E(y_m | \mathscr{L}_{m-1})| \leqq |E(v_m 1\{|v_m| > 50\|h\|_2 m\} | \mathscr{L}_{m-1})| + \|h\|_2 \\ &\leqq (50\|h\|_2 m^{-1}) E(v_m^2 | \mathscr{L}_{m-1}) + \|h\|_2 \leqq 2\|h\|_2 \log\log m. \end{aligned}$$

Lemma 5.5. *We have with probability* 1

$$n^{-1} \sum_{i<n} h(X_i, X_n) \to 0.$$

Proof. It is enough to show that for each $\varepsilon > 0$

$$n^{-1} \sum_{i<n} h(X_i, X_n) > \varepsilon \quad \text{only finitely often a.s.} \tag{5.6}$$

Since we can replace h by $19h/\varepsilon$ without violating (1.8) it suffices to show (5.6) with $\varepsilon = 19$. Put

$$\tau_n = \mathrm{n}/\log n$$

$$g_n(x,y) = h(x,y) 1\{h \leqq \tau_n\}$$

and

$$h_n(x,y) = g_n(x,y) - \int g_n(u,y)\,du - \int g_n(x,v)\,dv + \int\int g_n(u,v)\,du\,dv.$$

We shall prove that with probability 1 both

$$n^{-1} \sum_{i<n} (h(X_i, X_n) - h_n(X_i, X_n)) > 3 \tag{5.7}$$

and

$$n^{-1} \sum_{i<n} h_n(X_i, X_n) > 16 \tag{5.8}$$

happen only finitely often.

To prove (5.7) we set

$$I_1(n) = n^{-1} \sum_{i<n} h(X_i, X_n) 1\{h(X_i, X_n) > \tau_n\}$$

$$I_2(n) = n^{-1} \sum_{i<n} \int h(u, X_n) 1\{h(u, X_n) > \tau_n\}\,du$$

$$I_3(n) = n^{-1} \sum_{i<n} \int h(X_i, v) 1\{h(X_i, v) > \tau_n\}\,dv$$

and

$$I_4(n) = \int\int h(u,v) 1\{h > \tau_n\}\, du\, dv$$

and observe that the left side of (5.7) equals $\sum_{j=1}^{4} I_j(n)$. Now

$$(5.9) \qquad \begin{aligned} \sum_{n \geq 1} P(I_1(n) \neq 0) &\leq \sum_{n \geq 1} n P\{h(X_1, X_2) > \tau_n\} \\ &= \sum_{n \geq 1} n \sum_{j \geq n} P\{\tau_j < h(X_1, X_2) \leq \tau_{j+1}\} \\ &\ll \sum_{j \geq 1} j^2 P\{\tau_j < h(X_1, X_2) \leq \tau_{j+1}\} \\ &\ll \int h^2(x,y) \log h^2(x,y)\, dx\, dy < \infty. \end{aligned}$$

Next, since

$$\begin{aligned} I_2(n) &\leq \int |h(u, X_n)| 1\{|h(u, X_n)| > \tau_n\}\, du \\ &\leq \tau_n^{-1} \int h^2(u, X_n)\, du = \tau_n^{-1} A_n \qquad \text{(say)} \end{aligned}$$

we obtain by Jensen's inequality with $\phi(x) = x \log x$

$$(5.10) \qquad \begin{aligned} \sum_{n \geq 1} P(I_2(n) > 1) &\leq \sum_{n \geq 1} P(A_n > \tau_n) \leq \sum_{n \geq 1} P(\phi(A_n) > n) \\ &\ll E\phi(A_1) \ll E \int \phi(h^2(u, X_1))\, du < \infty. \end{aligned}$$

Further,

$$(5.11) \qquad \begin{aligned} \sum_{n \geq 1} P(I_3(n) > 1) &\leq \sum_{n \geq 1} n^{-1} E\left(\int h(X_i, v) 1\{h(X_i, v) > \tau_n\}\, dv\right)^2 \\ &\leq \sum_{n \geq 1} n^{-1} \int\int h^2(x,y) 1\{h(x,y) > \tau_n\}\, dx\, dy \\ &= \sum_{n \geq 1} n^{-1} \sum_{j \geq n} E h^2(X_1, X_2) 1\{\tau_j < h(X_1, X_2) \leq \tau_{j+1}\} \\ &\ll \sum_{j \geq 1} \log j\, E h^2(X_1, X_2) 1\{\tau_j < h(X_1, X_2) \leq \tau_{j+1}\} \\ &\ll E h^2(X_1, X_2) \log^2 h(X_1, X_2) < \infty. \end{aligned}$$

Since trivially $I_4(n) \to 0$ (5.7) follows from (5.9), (5.10), (5.11) and the Borel Cantelli lemma.

For the proof of (5.8) let X_0 be a random variable with $\mathscr{L}(X_0) = \mathscr{L}(X_1)$ and independent of the sequence $\{X_j, j \geq 1\}$. Let $\mathscr{F}_i$ be the σ-field generated by $X_0, X_1, \ldots, X_i$. Then $\{h_n(X_i, X_0), \mathscr{F}_i, 1 \leq i < n\}$ is a martingale difference sequence satisfying

$$(5.12) \qquad h_n(X_i, X_0) \leq c = 4\tau_n$$

and

$$(5.13) \qquad \begin{aligned} \sum_{i<n} E(h_n^2(X_i, X_0) | \mathscr{F}_{i-1}) &\leq 2 \sum_{i<n} \int h^2(u, X_0)\, du + 2 \sum_{i<n} \int\int h^2(x,y)\, dx\, dy \\ &< 2n \int h^2(u, X_0)\, du + 2n \int\int h^2(x,y)\, dx\, dy. \end{aligned}$$

We now apply Lemma 5.4.1 and Corollary 5.4.1 of Stout [21] with $\lambda=1/c$ and obtain

$$\begin{aligned}
&P\{n^{-1}\sum_{i<n} h_n(X_i,X_n)>16\}\\
&\quad=P\{\log n/(4n)\sum_{i<n} h_n(X_i,X_0)>4\log n\}\\
&\quad\leqq P\{\exp(\lambda\sum_{i<n} h_n(X_i,X_0))\cdot\exp(-\tfrac{1}{2}\lambda^2(1+\tfrac{1}{2}\lambda c)\\
&\qquad\cdot\sum_{i<n} E(h^2(X_i,X_0)|\mathscr{F}_{i-1}))>n^2\}\\
&\qquad+P\{\tfrac{1}{2}\lambda^2(1+\tfrac{1}{2}\lambda c)\sum_{i<n} E(h^2(X_i,X_0)|\mathscr{F}_{i-1})>2\log n\}\\
&\quad\leqq n^{-2}+P\{\int h^2(u,X_0)\,du>8\tau_n\}\\
&\quad\ll n^{-2}+P(A_n>8\tau_n)
\end{aligned}$$

by (5.13). Now (5.8) follows from (5.10) and the Borel Cantelli lemma. Since (5.7) and (5.8) together prove (5.6) with $\varepsilon=19$ Lemma 5.5 is proven. □

We now turn to the proof of Theorem 2. Several steps in the argument also will be used in the proof of Theorem 1, given at the end of this section. We modify the proof of Theorem 4 as given in Sect. 4. Let $\varepsilon_k\downarrow 0$ slowly and put

$$t_k=\prod_{j\leqq k}(1+\varepsilon_j),\qquad n_k=t_k-t_{k-1}=t_{k-1}\varepsilon_k. \tag{5.14}$$

Moreover, we choose sequences $M_k\uparrow\infty$ and $g_k\uparrow\infty$ both slowly at a rate to be determined later. Next, let f_l: $\mathbb{R}^2\to\mathbb{R}$, $l\geqq 1$ be a sequence of simple functions with sets of constancy being measurable rectangles $A_i\times A_j$ such that

$$\|h-f_l\|_2\leqq 2^{-l-1}/600. \tag{5.15}$$

Denote by $d(f_l)$ the number of values f_l assumes. By Proposition 5.1 we have for every $l\geqq 1$

$$\limsup_{n\to\infty}(n\log\log n)^{-1}V_n(h-f_l)\leqq 2^{-l-1}\qquad\text{a.s.}$$

and by Lemma 3.1 we have the same relation but with V_n replaced by W_n. Hence we can find a sequence $r_l\uparrow\infty$ such that

$$P\{\sup_{n\geqq r_l}(n\log\log n)^{-1}V_n(h-f_l)\geqq 2^{-l}\}\leqq 2^{-l}.$$

Thus by the Borel Cantelli lemma we have that

$$\sup_{n\geqq r_l}(n\log\log n)^{-1}V_n(h-f_l)\to 0\qquad\text{a.s.} \tag{5.16}$$

and

$$\sup_{n\geqq r_l}(n\log\log n)^{-1}W_n(h-f_l)\to 0\qquad\text{a.s.} \tag{5.17}$$

We now approximate h on $(t_k, t_{k+1}]$ by a subsequence of $\{f_l, l \geqq 1\}$ defined as follows. We let $l = l(k)$ be the largest integer satisfying

$$(5.18) \qquad r_l \leqq t_\kappa, \quad \|f_l\|_\infty \leqq M_k \quad \text{and} \quad d(f_l) \leqq d_k$$

where M_k, d_k and also ε_k tend to their respective limits slowly at a rate still to be determined. We define

$$(5.19) \qquad g_k = f_{l(k)}.$$

Since $l(k) \to \infty$ we have $\|h - g_k\|_2 \to 0$.

We now follow the proof of Theorem 3. Relation (4.9) becomes

$$\pi(\mathscr{L}(n_k^{-1} R_k), \mathscr{N}(O, C_k)) \ll n_k^{-\frac{1}{9}} d_k^{\frac{1}{3}}.$$

We can assume without loss of generality that $\varepsilon_k \downarrow 0$ and $d_k \uparrow \infty$ so slowly that $\sum_{k \geqq 1} n_k^{-\frac{1}{90}} d_k < \infty$. Then (4.12) gets replaced by

$$(5.20) \qquad \Big(\sum_{i \leqq d_k} (R(A(i,k), H_k) - K(A(i,k), H_k))^2\Big)^{\frac{1}{2}} \ll n_k^{\frac{1}{2}} n_k^{-\frac{1}{9}} d_k^{\frac{1}{3}} \ll n_k^{\frac{4}{10}}.$$

Lemma 5.6. *Let $I_k = (t_\kappa, t_k)$. Then we have with probability* 1

$$\int\int g_\kappa(x, y)(R(dx, I_k) R(dy, I_k) - K(dx, I_k) K(dy, I_k)) \ll n_k.$$

Proof. We follow the proof of Lemma 4.2. The bound in (4.19) is replaced by $d_k^{\frac{1}{2}} n_k^{\frac{4}{10}}$. (4.20) still reads $\|K_l\| \ll n_l^{\frac{1}{2}} l^{\frac{1}{2}}$. Hence the bound in (4.21) becomes $M_k d_k^{\frac{1}{2}} n_k^{\frac{4}{10}} n_k^{\frac{1}{2}} d_k^{\frac{1}{2}} k^{\frac{1}{2}}$. Thus we obtain in the lemma the bound $M_k d_k k^{\frac{5}{2}} n_k^{\frac{9}{10}} \ll n_k$ if both $\varepsilon_k \downarrow 0$ and $M_k \uparrow \infty$ sufficiently slowly. □

The following lemma is an immediate extension of Lemma 5.4.1 and Corollary 5.4.1 of Stout [21].

Lemma 5.7. *Let $\{U_n, \mathscr{F}_n, n \geqq 1\}$ be a supermartingale with $EU_1 = 0$. Let $U_0 = 0$ and $Y_i = U_i - U_{i-1}$ for $i \geqq 1$. Suppose $Y_i \leqq c$ a.s. for some $0 \leqq c < \infty$ and all $i \geqq 1$. Let $\lambda > 0$ and*

$$T_n = \exp(\lambda U_n) \exp\Big(-\tfrac{1}{2}\lambda^2 e^{\lambda c} \sum_{i \leqq n} E(Y_i^2 | \mathscr{F}_{i-1})\Big), \qquad n \geqq 1$$

and $T_0 = 1$ a.s. Then $\{T_n, \mathscr{F}_n, n \geqq 0\}$ is a non-negative supermartingale and for each $\alpha > 0$

$$P\{\sup_{n \geqq 0} T_n > \alpha\} < 1/\alpha.$$

Lemma 5.8. *As $k \to \infty$ we have with probability* 1

$$\max_{t_k < n \leqq t_{k+1}} |V_n(h) - V_{t_k}(h)| = o(t_k \log\log t_k).$$

Proof. We use the notation introduced earlier in this section. Because of Lemmas 5.4 and 5.5 it is enough to show that

$$(5.21) \qquad \max_{t_k < n \leqq t_{k+1}} \Big| \sum_{m = t_k + 1}^{n} w_m \Big| = o(t_k \log\log t_k) \quad \text{a.s.}$$

To prove this we apply Lemma 5.7 to the martingale difference sequence $\{w_m, \mathscr{L}_m, t_k < m \leqq t_{k+1}\}$ with $c = 50\,\|h\|_2\, t_{k+1}$ and $\lambda = 1/(c\phi(k))$ where $\phi(k)\downarrow 0$ is chosen such that

$$e^{1/\phi(k)}\varepsilon_k/\phi^2(k) \to 0. \tag{5.22}$$

We obtain with $\alpha = \exp(4\log\log t_k)$

$$P\left\{\max_{t_k < n \leqq t_{k+1}} \exp\left(\lambda \sum_{m=t_k+1}^{n} w_m\right) \exp\left(-\tfrac{1}{2}\lambda^2 e^{\lambda c} \sum_{m=t_k+1}^{n} E(w_m^2 | \mathscr{L}_{m-1})\right) > \alpha\right\}$$
$$\leqq 1/\alpha \ll k^{-2}$$

if $\varepsilon_k \downarrow 0$ so slowly that $t_k \geqq \exp k^{\frac{1}{2}}$. Hence by Lemma 5.2 and (5.14) we obtain with probability 1 a $k_0(\omega)$ such that for all $k \geqq k_0$

$$\max_{t_k < n \leqq t_{k+1}} \lambda \sum_{m=t_k+1}^{n} w_m \leqq 4\log\log t_k + \tfrac{1}{2}\lambda^2 e^{\lambda c} \sum_{m=t_k+1}^{t_{k+1}} E(w_m^2 | \mathscr{L}_{m-1})$$
$$\ll (4 + e^{1/\phi(k)}\varepsilon_k/(5{,}000\,\phi^2(k)))\log\log t_k.$$

Substituting λ and using (5.22) we obtain (5.21) and thus the lemma. □

Lemma 5.9. *As $k \to \infty$ we have with probability* 1

$$\int\int (h(x,y) - g_\kappa(x,y))\, R(dx, I_k)\, R(dy, I_k) = o(t_k \log\log t_k).$$

Proof. Let $\varepsilon > 0$. By (5.16) there is a set Ω_0 with $P(\Omega_0) = 1$ and an $l_0 = l_0(\varepsilon, \omega)$ such that for all $\omega \in \Omega_0$ and $l \geqq l_0$

$$\sup_{n \geqq r_l} (n \log\log n)^{-1}\, V_n(h - f_l) \leqq \varepsilon.$$

Let k be so large that $l(\kappa) \geqq l_0$, so $g_\kappa = f_l$ for some $l \geqq l_0$. Moreover, by (5.18) $t_k \geqq r_{l(k)} \geqq r_{l(\kappa)}$. Hence for all $\omega \in \Omega_0$

$$(t_k \log\log t_k)^{-1}\, V_{t_k}(h - g_\kappa) \leqq \sup_{n \geqq r_{l(\kappa)}} (n\log\log n)^{-1}\, V_n(h - f_{l(\kappa)}) \leqq \varepsilon.$$

This shows that

$$V_{t_k}(h - g_\kappa) = o(t_k \log\log t_k).$$

Since Lemma 4.4 remains valid if $\varepsilon_k \downarrow 0$ so slowly that $t_k \geqq \exp k^{\frac{3}{4}}$ and with h replaced by $h - g$ this last relation implies the lemma. □

We need Lemmas 5.8 and 5.9 but with V and R replaced by W and K respectively.

Lemma 5.10. *As $k \to \infty$ we have with probability* 1

$$\max_{t_k < n \leqq t_{k+1}} |W_n(h) - W_{t_k}(h)| = o(t_k \log\log t_k).$$

Proof. As in the proof of Lemma 3.1 we have

$$
\begin{aligned}
&P\left\{\max_{t_k<n\leqq t_{k+1}} \sum_{m=t_k+1}^{n} (w_m-c)\geqq 20\|h\|_2 n_k \log\log t_k\right\}\\
&\quad=P\left\{\max_{t_k<n\leqq t_{k+1}} \exp\left(n_k^{-1} t \sum_{m=t_k+1}^{n} (w_m-c)\right)\geqq \exp(4\log\log t_k)\right\}\\
&\quad\leqq \exp(-4\log\log t_k) E\exp\left(n_k^{-1} t \sum_{m=t_k+1}^{t_{k+1}} w_m\right)\exp(-ct)\\
&\quad\ll k^{-2} E\exp(t W_1(h))\ll k^{-2}.
\end{aligned}
$$

The lemma follows now from the Borel Cantelli lemma since $n_k=\varepsilon_k t_{k-1}$.

Lemma 5.11. *As* $k\to\infty$ *we have with probability* 1

$$\iint (h(x,y)-g_\kappa(x,y)) K(dx, I_k) K(dy, I_k)=o(t_k \log\log t_k).$$

Proof. Since the analogue of Lemma 4.4 remains valid with R replaced by K it is enough to prove that with probability 1

$$W_{t_k}(h-g_\kappa)=o(t_k\log\log t_k).$$

But this follows from Theorem 5 in much the same way as Lemma 3.1. We have with $t=(5\|h-g_\kappa\|_2)^{-1}$ and $c=E\iint (h-g_\kappa)(x,y) B_1(dx) B_1(dy)$

$$
\begin{aligned}
&P\{\max_{n\leqq t_k}(W_n(h-g_\kappa)-nc)\geqq 20 t_k\|h-g_\kappa\|_2 \log\log t_k\}\\
&\quad\leqq \exp(-4\log\log t_k) E\exp(t t_k^{-1}(W_{t_k}(h-g_\kappa)-t_k c))\\
&\quad\ll k^{-2} E\exp(t W_1(h-g_\kappa))\ll k^{-2}. \quad \square
\end{aligned}
$$

The proof of Theorem 2 can now be completed as in the last paragraph of Sect. 4.

For the proof of Theorem 1 we replace Lemma 5.8 and 5.9 by the following one and observe that they remain valid if we replace V and R by W and K respectively.

Lemma 5.12. *As* $k\to\infty$

$$t_k^{-1}\max_{t_k<n\leqq t_{k+1}} |V_n(h)-V_{t_k}(h)|\to 0 \quad \text{in probability.}$$

$$t_k^{-1}\iint (h(x,y)-g_\kappa(x,y)) R(dx, I_k) R(dy, I_k)\to 0 \quad \text{in probability.}$$

Proof. This is an immediate consequence of Doob's generalization of Kolmogorov's inequality, of Lemma 2.1 and since $\varepsilon_k\downarrow 0$. Recall that as was noted in the proof of Lemma 2.4 $\{V_n(h)-n\int h(x,x)dx, \mathscr{L}_n, n\geqq 1\}$ is a martingale.

6. Extensions

Theorem 4 can easily be extended to kernels in more than two arguments and to the multivariate case. Let $\{X_j, j\geqq 1\}$ be a sequence of independent identically distributed random vectors in $\mathbb{R}^q$, $q \geqq 1$ and let $\{R(s,t),\ s \in \mathbb{R}^q,\ t \geqq 0\}$ be the empirical process of $\{X_j, j \geqq 1\}$. Let $h\colon \mathbb{R}^{mq} \to \mathbb{R}^d$ be a measurable function; here $m \geqq 2$. Suppose that for some $\delta \geqq 0$

$$(6.1) \qquad \int_{\mathbb{R}^{mq}} |\hat{h}(y_1, \ldots, y_s)|^{2+\delta} dF(y_1) \ldots dF(y_s) < \infty \qquad s = 1, 2, \ldots, m$$

with the following interpretation. F is the common distribution function of $\{X_j, j \geqq 1\}$, and $\hat{h}$ is defined by a fixed partition $C_1, \ldots, C_s$ of $\{1, \ldots, m\}$ via the relation $\hat{h}(y_1, \ldots, y_s) = h(x_1, \ldots, x_m)$ with $x_i = y_j$ iff $i \in C_j$. Then $\|h\|_{2+\delta}$ is defined as the sum over all possible integrals of the form (6.1) raised to the power $1/(2+\delta)$. For $\|h\|_2 < \infty$ the von Mises statistic is defined as the $\mathbb{R}^d$-valued process

$$V_n(h) = \int h(x_1, \ldots, x_m) R(dx_1, n) \ldots R(dx_m, n).$$

The Kiefer process figuring in the approximating integrals has covariance function $EK(s,t)K(s',t) = \min(t,t')(F(s \wedge s') - F(s) \cdot F(s'))$, $s, s' \in \mathbb{R}^q$. Here $s \wedge s'$ is the vector with components being the minimum of the corresponding components. It is clear now how to reformulate Theorem 4 to conform with this more general situation. However, in the error terms the exponent 1 on n has to be replaced by $\frac{1}{2}m$. The changes required in the proof are routine.

The extension to the multisample case, i.e. to statistics of the form

$$\iint h(x_1, \ldots, x_{m_1}, y_1, \ldots, y_{m_2}) \prod_{i \leqq m_1} R(dx_i, n_1) \prod_{j \leqq m_2} R^*(dy_j, n_2)$$

where R and R^* are the empirical processes of independent samples appears not to be obvious. If $n_1 = n_2$ the methods of the present paper still work with virtually no changes. However, if $n_1 = n_2$ is not assumed then presumably the methods of [1] combined with the methods of the present paper will lead to the desired extension.

7. Hölder Continuity and Bounded Variation

In this section we give two sufficient conditions on h which guarantee that (1.10) holds. All the standard kernels satisfy one of these conditions. A function $h\colon \mathbb{R}^m \to \mathbb{R}$ is called Hölder continuous with exponents ρ, $r > 0$ and constant C if for all $x_1, \ldots, x_m, y_1, \ldots, y_m$

$$(7.1) \qquad |h(x_1, \ldots, x_m) - h(y_1, \ldots, y_m)| \leqq C \sum_{i,j \leqq m} |x_i - y_i|^\rho (1 + |x_j|^r + |y_j|^r).$$

Lemma 7.1. *Let $h\colon \mathbb{R}^m \to \mathbb{R}$ be Hölder continuous with exponents ρ, $r > 0$ and constant C and suppose that F is a distribution function on $\mathbb{R}^m$ having a moment of order $2(r+\rho)+\delta$ for some $\delta > 0$. Then there is a constant D with the following property: For every $d \in \mathbb{N}$ there exists a partition $\alpha = \{A(i,d),\ 1 \leqq i \leqq d\}$ of $\mathbb{R}$ and $g(i_1, \ldots, i_m) \in \mathbb{R}$ such that*

$$\|h - \sum_{i_1, \ldots, i_m} g(i_1, \ldots, i_m) 1_{A(i_1,d) \times \ldots \times A(i_m,d)}\|_2 \leqq D d^{-\rho\delta/(2\rho+\delta)}.$$

Proof. Recall that $\|h\|_2$ was defined in Sect. 6. Put

$$(7.2)\qquad K=d^{2\rho/(2\rho+\delta)}, \quad M=\tfrac{1}{2}K^{\frac{1}{2}\delta/\rho},$$
$$c_i=-K+(i-1)/M, \quad 1\leqq i\leqq d, \quad c_0=-\infty, \quad c_d=+\infty.$$

For $0\leqq i_j<d$ $(1\leqq j\leqq m)$ we set

$$(7.3)\qquad g(i_1,\ldots,i_m)=h(c_{i_1},\ldots,c_{i_m})$$
$$A(i_j,d)=[c_{i_j},c_{i_j+1})$$

The lemma follows easily by elementary calculations.

Corollary 7.2. *Under the hypotheses of Lemma* 7.1 *condition* (1.10) *is satisfied with* $\gamma=\rho\delta/(2\rho+\delta)$.

Proof. For $r=1,2,\ldots$ apply Lemma 7.1 with $d=2^r$. As can be seen from the proof of Lemma 7.1 we obtain for each $r=1,2,\ldots$ a sequence $-\infty=c_0^r<c_1^r<\ldots<c_{2^r}^r=+\infty$ and simple functions

$$h^r=\sum_{0\leqq i_j<2^r} g^r(i_1,\ldots,i_m)1_{A(i_1,2^r)\times\ldots\times A(i_m,2^r)}$$

where $A(i,2^r)=[c_i^r,c_{i+1}^r)$, $0\leqq i<2^r$ such that $\|h-h^r\|_2\leqq C_0 2^{-r\gamma}$. We define a new partition $\beta(r)$ of $\mathbb{R}$ into (at most) 2^r sets $\{-\infty=b_0\leqq b_1\leqq\ldots\leqq b_{2^r}=\infty\}$ by reordering $\bigcup_{s<r}\bigcup_{1\leqq i<2^s}\{c_i^s\}$. $\{\beta(r), r\geqq 2\}$ is a refining sequence of partitions and h^{r-1} is measurable with respect to the σ-field generated by $\beta(r)$. □

The second condition is on the mean oscillation of $h\colon[0,1]^m\to\mathbb{R}$. For $\varepsilon>0$ and $x\in\mathbb{R}^m$ put

$$\operatorname{osc}(h,\varepsilon,x)=\sup\{|h(x)-h(y)|\colon y\in\mathbb{R}^m, |x-y|<\varepsilon\}.$$

Lemma 7.3. *Suppose that* $h\colon[0,1]^m\to\mathbb{R}$ *satisfies*

$$(7.4)\qquad \sup_{\varepsilon>0}\varepsilon^{-r}\int_{\mathbb{R}^m}(\operatorname{osc}(h,\varepsilon,x))^2\,dx<\infty$$

for some $r>0$. *Moreover suppose that* (7.4) *holds along each diagonal of* $[0,1]^m$. *Then the conclusions of Lemma* 7.1 *and Corollary* 7.2 *hold with exponent* $-\rho\delta/(2\rho+\delta)$ *replaced by* $-r$.

Proof. Let $c_i=i/d$, $0\leqq i\leqq d$ and define $g(i_1,\ldots,i_m)$ and $A(i,d)$ by (7.3). Then for each $x\in A(i_1,d)\times\ldots\times A(i_m,d)$

$$|h(x)-g(i_1,\ldots,i_m)|\leqq\operatorname{osc}(h,1/d,x).\quad\square$$

8. Examples

Theorems 1 through 4 immediately apply to the standard examples frequently mentioned in this context. (See [10].)

8.1. For the estimator of the sample variance S^2 we have the relation

$$n^2S^2=n\sum_{i\leqq n}(X_i-\bar{X})^2=\sum_{1\leqq i,j\leqq n}(X_i-X_j)^2=\sum_{1\leqq i\neq j\leqq n}(X_i-X_j)^2.$$

So we can set $m=2$ and $h(x,y)=\frac{1}{2}(x-y)^2$. Of course, h is non-degenerate, but setting $h_1(x)=\frac{1}{2}x^2-x\int y\,dF(y)+\frac{1}{2}\int y^2\,dF(y)$ and $\sigma^2=\iint h(x,y)\,dF(x)\,dF(y)$ we have

$$n^2(S^2-2\sigma^2)=2V_n(h)+4n^2\int h_1(x)\,d(F_n(x)-F(x)).$$

Assuming suitable moment conditions on F we conclude from (1.13), (1.14), etc. that $n^{\frac{1}{2}}(S^2-2\sigma^2)$ can be approximated by the same Brownian motion as $4n^{\frac{1}{2}}\int h_1(x)\,d(F_n(x)-F(x))$, a result due to Sen [20]. However, according to Theorems 2 or 3 together with Lemma 7.1 and Corollary 7.2 the difference of these two expressions can be approximated by a suitable W_n. In particular, we obtain from Theorems 3 and 2 with probability 1

$$\limsup_{n\to\infty} n^{\frac{1}{2}}/(2\log\log n)(n^{\frac{1}{2}}(S^2-2\sigma^2)-4n^{\frac{1}{2}}\int h_1(x)\,d(F_n(x)-F(x)))=2C(h^*)$$

where $C(h^*)=\sigma^2$.

8.2. Similarly for the Wilcoxon signed rank statistic the function

$$\begin{aligned} h(x,y)&=1 \quad \text{if } x+y>0\\ &=0 \quad \text{otherwise}\end{aligned}$$

is non-degenerate and one can proceed as in Sect. 8.1, replacing Lemma 7.1 and Corollary 7.2 by Lemma 7.3.

8.3. In the test for the sample covariance the function h we are interested in is $h((x_1,y_1),(x_2,y_2))=x_1y_1-x_1y_2$ since

$$\begin{aligned}\frac{1}{n}\sum_{i\le n}(X_i-\bar{X})(Y_i-\bar{Y})&=\frac{1}{n}\sum_{i\le n}X_iY_i-\bar{X}\bar{Y}\\ &=\int(x_1y_1-x_1y_2)\,dF_n(x_1,y_1)\,dF_n(x_2,y_2).\end{aligned}$$

So $m=q=2$ and the results of Sect. 6 extending Theorem 4 apply.

8.4. Let F be a distribution function and $I_1,\ldots,I_L$ be a partition of the real line with probabilities $F(I_l)=p_l$, $1\le l\le L$. Let $\{X_i, i\ge 1\}$ be a sequence of independent random variables with common distribution function F. Put

$$h(x,y)=\sum_{l\le L}p_l^{-1}(1_{I_l}(x)-p_l)(1_{I_l}(y)-p_l).$$

Then the statistic

$$\chi(n,L)=\sum_{l\le L}p_l^{-1}(\operatorname{card}\{j\le n: X_j\in I_l\}-np_l)^2=\sum_{1\le i,j\le n}h(X_i,X_j)$$

is just the one figuring in the χ^2-test of fit. Notice that h is degenerate as $\int h(x,y)\,dF(x)=0$. The function h satisfies the conditions of Theorem 4. Hence for some Kiefer process and some $\lambda>0$

$$\chi(n,L)-\iint h(x,y)\,K(dx,n)\,K(dy,n)\ll n^{1-\lambda} \quad \text{a.s.}$$

One can easily derive $C(h^*)=1$ (cf. [9]).

8.5. Let $\alpha(n)=\sqrt{n}\,\frac{F_n(x)-x}{x(1-x)}$ $(0<x<1)$ and consider the statistic

$$nA_n=\int_0^1 \alpha_n^2(x)\,dx,$$

where F_n denotes the empirical distribution function of a sample of size n taken from the uniform distribution on $[0,1]$. It is easy to see that $nA_n = n^{-1} \sum_{1\leqq i,j\leqq n} h(X_i, X_j)$ is derived from a von Mises' functional with kernel

$$h(x,y)=\int_0^1 u(1-u))^{-1}(1\{x\leqq u\}-u)(1\{y\leqq u\}-u)\,du.$$

From Corollary 1 and the result of de Wet and Venter [5] we deduce $\limsup_{n\to\infty} (\log\log n)^{-1} nA_n=1$. This result has been obtained by Csáki [3, Theorem 3.1].

References

1. Berkes, I., Morrow, G.J.: Strong invariance principles for mixing random field. Z. Wahrscheinlichkeitstheorie verw. Gebiete **57**, 15–37 (1981)
2. Berkes, I., Philipp, W.: Approximation theorems for independent and weakly dependent random vectors. Annals Probability **7**, 29–54 (1979)
3. Csáki, E.: On the standardized empirical distribution function. Coll. Mathem. Soc. János Bolyai, **32**. Nonparametric Statistical Inference, pp. 123–138. Budapest (1980)
4. Denker, M., Grillenberger, C., Keller, G.: A note on invariance principles for v. Mises' statistics. Metrika, to appear
5. de Wet, T., Venter, J.H.: Asymptotic distributions for quadratic forms with applications to tests of fit. Ann. Statist. **1**, 380–387 (1973)
6. Doob, J.L.: Stochastic Processes. New York: Wiley 1953
7. Fernique, X.: Intégrabilité des vecteurs gaussiens. C.R. Acad. Sci. Paris Sér. A **270**, A1698–1699 (1970)
8. Fillipova, A.A.: Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. Theory Probability Applications **7**, 24–57 (1962)
9. Gregory, G.G.: Large sample theory for *U*-statistics and tests of fit. Ann. Statist. **5**, 110–123 (1977)
10. Hall, P.: On the invariance principle for *U*-statistics. Stochastic Processes Applications **9**, 163–174 (1979)
11. Hoeffding, W.: A class of statistics with asymptotically normal distribution. Ann. Math. Statist. **19**, 293–325 (1948)
12. Kuelbs, J.: Kolmogorov's law of the iterated logarithm for Banach space valued random variables. Illinois J. Math. **21**, 784–800 (1977)
13. Loynes, R.M.: On the weak convergence of *U*-statistic processes and of the empirical processes. Math. Proc. Cambridge Philos. Soc. **83**, 269–272 (1978)
14. Miller, R.G. Jr., Sen, P.K.: Weak convergence of *U*-statistics and von Mises' differentiable statistical functions. Ann. Math. Statist. **43**, 31–41 (1972)
15. von Mises, R.: On the asymptotic distribution of differentiable statistical functions. Ann. Math. Statist. **18**, 309–348 (1947)
16. Moricz, F.: Moment inequalities for the maximum of partial sums of random fields. Acta Sci. Math. Hungar. **39**, 353–366 (1977)

17. Morrow, G., Philipp, W.: An almost sure invariance principle for Hilbert space valued martingales. Trans. Amer. Math. Soc. **273**, 231–251 (1982)
18. Neuhaus, G.: Functional limit theorems for U-statistics in the degenerate case. J. Multivariate Analysis **7**, 424–439 (1977)
19. Philipp, W.: Almost sure invariance principles for sums of B-valued random variables, Probability in Banach spaces, II, (Proc. Conf. Oberwolfach, 1978). Lecture Notes in Math. **709**, 171–193. Berlin-Heidelberg-New York: Springer 1979
20. Sen, P.K.: Almost sure behaviour of U-statistics and von Mises' differentiable statistical functions. Ann. Statist. **2**, 387–396 (1974)
21. Stout, W.F.: Almost Sure Convergence. New York: Academic Press 1974
22. Yurinskii, V.V.: On the error of the Gaussian approximation for convolutions. Theory Probability Appl. **22**, 236–247 (1977)

Received July 20, 1983; in revised form March 30, 1984