JOURNAL OF MULTIVARIATE ANALYSIS 16, 335–367 (1985)

Multilinear Forms and Measures of Dependence between Random Variables

RICHARD C. BRADLEY*

Indiana University

AND

WLODZIMIERZ BRYC

Instytut Matematyki, Politechnika Warszawska, Poland

Communicated by L. D. Pitt

Based on an idea of Rosenblatt, the methods of interpolation theory are used to establish moment inequalities and equivalence relations for measures of dependence between two or more families of random variables. A couple of "interpolation" theorems proved here appear to be new.

I. INTRODUCTION

In his studies of mixing conditions on Markov chains, Rosenblatt [32; 33, Chap. 7] used the Riesz convexity (interpolation) theorem to compare different measures of dependence between two given families of random variables on a probability space. Rosenblatt [34] also suggested that by using other results in operator theory, one might be able to obtain more information about the relationships between various measures of dependence. In this article we shall follow up this suggestion and, in essence, see what more information can be obtained from the Riesz–Thorin and Marcinkiewicz interpolation theorems and from a key idea of Stein and Weiss [37].

The nature of this paper is partly expository, pointing out relevant

Received January 16, 1984.

AMS subject classification: Primary 60B05; Secondary 46M35.

Key words and phrases: Measures of dependence, multilinear forms, multilinear operators, moment inequalities, equivalence relation, interpolation.

* The work of R.C.B. was supported in part by NSF grant MCS 81-01583.

0047-259X/85 $3.00

applications of some theorems that are either elementary or well known to functional analysts. This first section is a brief introduction, to serve as motivation. Sections 2 and 3 give a discussion of the relevant results in interpolation theory on spaces of functions; Theorems 2.1 and 3.6 there appear to be new and may be of independent interest in interpolation theory. In Section 4 we return to the context of measures of dependence on a probability space, and apply the results in Sections 2–3 to that context.

This paper developed from the authors' work in the following way. Theorem 4.3 and Example 4.4 came (essentially verbatim) from an earlier, unpublished manuscript of R.C.B. After seeing that work, [33, Chap. 7], and preprints of [3] and [28], W.B. spotted potential broad applicability of interpolation theory to measures of dependence and proved Theorem 2.1 for the case $n=2$ (including Theorem 2.2), Theorem 3.6 for the case $n=2$, and Theorem 1.1, and (for expository purposes) worked out Theorem 4.2 in Section 4.3. Then the present (multidimensional) versions of Theorems 2.1 and 3.6, along with other odds and ends, were worked out jointly.

Let $(\Omega, \mathscr{M}, P)$ be a probability space. Two sub-σ-fields $\mathscr{F}$ and $\mathscr{G} \subset \mathscr{M}$ are said to be "independent" if $P(A \cap B) = P(A)\, P(B)$, $\forall\, A \in \mathscr{F}$, $B \in \mathscr{G}$. This definition is the starting point for the following class of measures of dependence between σ-fields: For $0 \leqslant r \leqslant 1$, $0 \leqslant s \leqslant 1$, and any two σ-fields $\mathscr{F}$ and $\mathscr{G}$ $(\subset \mathscr{M})$ define

$$\alpha_{r,s}(\mathscr{F}, \mathscr{G}) := \sup \frac{|P(A \cap B) - P(A)\, P(B)|}{[P(A)]^r [P(B)]^s},$$
$$A \in \mathscr{F},\ B \in \mathscr{G},\ P(A) > 0,\ P(B) > 0. \tag{1.1}$$

For certain ordered pairs (r, s) these measures of dependence are already known, by the following notations:

$$\begin{aligned}
\alpha(\mathscr{F}, \mathscr{G}) &:= \alpha_{0,0}(\mathscr{F}, \mathscr{G}),\\
\phi(\mathscr{F}, \mathscr{G}) &:= \alpha_{1,0}(\mathscr{F}, \mathscr{G}).\\
\phi(\mathscr{G}, \mathscr{F}) &= \alpha_{0,1}(\mathscr{F}, \mathscr{G}),\\
\psi(\mathscr{F}, \mathscr{G}) &:= \alpha_{1,1}(\mathscr{F}, \mathscr{G}),\\
\lambda(\mathscr{F}, \mathscr{G}) &:= \alpha_{1/2,1/2}(\mathscr{F}, \mathscr{G}).
\end{aligned} \tag{1.2}$$

The quantities $\alpha(\cdot, \cdot)$, $\phi(\cdot, \cdot)$, and $\psi(\cdot, \cdot)$ are the measures of dependence used respectively in the "strong mixing," "ϕ-mixing," and "ψ-mixing" conditions for sequences of random variables. For the definitions of these conditions, see, e.g. [21] or [22]. The use of these mixing conditions in central limit theory started with Rosenblatt's [31] use of $\alpha(\cdot,\cdot)$ (in the "strong mixing" condition).

Throughout this article, the random variables will be complex-valued. For $1 \leqslant p \leqslant \infty$, let $\|X\|_p$ denote the usual p-norm of any given r.v. X (i.e., $\|X\|_p = E^{1/p}|X|^p$ if $1 \leqslant p < \infty$, and $\|X\|_\infty = P$-ess sup $|X|$), and for any such p and any σ-field $\mathscr{F}$ ($\subset \mathscr{M}$) let $\mathscr{L}_p(\mathscr{F})$ denote the class of (complex-valued) $\mathscr{F}$-measurable r.v.'s X satisfying $\|X\|_p < \infty$.

The following theorem is given here in order to help focus our discussion:

THEOREM 1.1. *Suppose* $1 \leqslant p, q \leqslant \infty$ *and* $p^{-1} + q^{-1} \leqslant 1$. *Suppose* $\mathscr{F}$ *and* $\mathscr{G}$ *are* σ-*fields*, $X \in \mathscr{L}_p(\mathscr{F})$, *and* $Y \in \mathscr{L}_q(\mathscr{G})$. *Then the following two statements hold:*

(i) *Defining* t, $1 \leqslant t \leqslant \infty$, *so that* $p^{-1} + q^{-1} + t^{-1} = 1$, *one has*

$$|EXY - EXEY| \leqslant 2\pi \cdot [\alpha(\mathscr{F}, \mathscr{G})]^{1/t} \cdot [\phi(\mathscr{F}, \mathscr{G})]^{1/p} \cdot [\phi(\mathscr{G}, \mathscr{F})]^{1/q} \cdot \|X\|_p \cdot \|Y\|_q.$$

(ii) *If* $p^{-1} + q^{-1} = 1$, *then*

$$|EXY - EXEY| \leqslant 3000 \cdot (\lambda(\mathscr{F}, \mathscr{G}) \cdot [1 - \log \lambda(\mathscr{F}, \mathscr{G})])^{\min(2/p, 2/q)} \cdot \|X\|_p \cdot \|Y\|_q.$$

Here and in what follows, log always denotes the natural logarithm. Throughout this paper, when $\log \alpha$ appears for some positive number α, it will turn out to be the case that $\alpha \leqslant 1$ and hence $1 - \log \alpha \geqslant 1$. Theorem 1.1 will be proved at the end of Section 4.1. In Section 4.3 it will be extended (with minor adjustments) to r.v.'s X and Y taking their values in a Hilbert space, following an idea in [8]. In Example 4.4 in Section 4.4 it will be shown that the log term in (ii) cannot be entirely avoided. Throughout this paper, except in Sections 2.2 and 4.4, "large" multiplicative constants will be permitted for the sake of keeping the proofs simple.

Part (i) of Theorem 1.1 was motivated partly by a preprint of Peligrad [28], in which part (i) was proved for the case $p^{-1} + q^{-1} = 1$. In both [9] and [28] there are limit theorems involving ϕ-mixing in both directions of time simultaneously. Except for a constant factor, (i) gives a unified treatment of two different families of moment inequalities (one involving $\alpha(\cdot,\cdot)$ and the other involving $\phi(\cdot,\cdot)$) that have been discussed in [5–7, 10, 12, 17, 18, 20, 21, 28, 38].

Part (ii) sharpens and generalizes an equivalence relation proved in [3] between $\lambda(\mathscr{F}, \mathscr{G})$ and the well known "maximal correlation" [13, 16]:

$$\rho(\mathscr{F}, \mathscr{G}) := \sup \frac{|EXY - EXEY|}{\|X\|_2 \cdot \|Y\|_2}, \qquad X \in \mathscr{L}_2(\mathscr{F}),\ Y \in \mathscr{L}_2(\mathscr{G}),$$

$$\|X\|_2 > 0,\ \|Y\|_2 > 0$$

$$= \sup |\mathrm{Corr}(X, Y)|, \qquad X \in \mathscr{L}_2(\mathscr{F}),\ Y \in \mathscr{L}_2(\mathscr{G}),\ X, Y \text{ real}. \quad (1.3)$$

(The latter equality is well known (see [39, p. 512, Theorem 1.1]); keep in mind the trivial fact that $\|W\|_2 \geqslant \|W - EW\|_2$ for $W \in \mathscr{L}_2(\mathscr{M})$.) Let us mention one application of (ii) to stochastic processes. Suppose $(X_k, k = ..., -1, 0, 1, ...)$ is a strictly stationary sequence of *real*-valued r.v.'s with $EX_k = 0$ and $EX_k^2 < \infty$. For each $n = 1, 2, ...$ define $\lambda(n) := \lambda(\mathscr{F}_{-\infty}^0, \mathscr{F}_n^\infty)$, where $\mathscr{F}_J^L$ denotes the σ-field of events generated by $(X_k, J \leqslant k \leqslant L)$. Then by (ii) the following statement is an immediate corollary of a theorem in a paper by Ibragimov [19]:

COROLLARY 1.2 (of [19, Theorem 2.2]). *If (X_k) satisfies $\lambda(n) = O((\log n)^{-(1+\varepsilon)})$ as $n \to \infty$ for some $\varepsilon > 0$, then (X_k) has a continuous spectral density $f(\lambda)$, and if in addition $f(0) \neq 0$ then $(X_1 + \cdots + X_n)/[2\pi n \cdot f(0)]^{1/2} \to N(0, 1)$ in distribution as $n \to \infty$.*

Remark 1.1. The log term that occurs in Theorem 1.1(ii) may turn out to be quite prevalent in moment inequalities. Consider the following result of Zuev [40]: If X is real and $\mathscr{F}$-measurable, Y is real and $\mathscr{G}$-measurable, $E\exp(a|X|) \leqslant C$ and $E\exp(a|Y|) \leqslant C$, where $a > 0$ and $C > 0$ are constants, then $|EXY - EXEY| \leqslant 8a^{-2}C \cdot \phi(\mathscr{F}, \mathscr{G}) \cdot [1 - \log \phi(\mathscr{F}, \mathscr{G})]$; and this inequality is sharp up to a multiplicative factor that depends only on C. Thus, in the absence of further information, the log term is in essence unavoidable. Žurbenko and Zuev [42] obtained earlier a very similar inequality, $|EXY - EXEY| \leqslant 72a^{-2}C \cdot \alpha(\mathscr{F}, \mathscr{G}) \cdot [\log \alpha(\mathscr{F}, \mathscr{G})]^2$, under the same conditions on X and Y.

In Section 4 our discussion on measures of dependence will be continued. In Section 4.2, measures of dependence between three or more σ-fields will be examined.

In what follows, when a term like a_b is to be a subscript or exponent, it will usually be written as $a(b)$ for typographical convenience. Vectors will be denoted with bold-face type: thus $\mathbf{p}$ denotes a vector and p a scalar. For any positive integer J, $[1, \infty]^J$ will denote the set of all vectors $\mathbf{p} = (p_1, ..., p_J)$ such that $1 \leqslant p_j \leqslant \infty\ \forall j = 1, ..., J$.

II. Multilinear Operators

This section is devoted mainly to Theorem 2.1 below, which is closely related to, and is based heavily on the ideas in, the Marcinkiewicz interpolation theorem. Theorem 2.1 is (ultimately) the basis for the proof of Theorem 1.1(ii), and (as we shall note in Section 4.2 later on) can (ultimately) be used in the study of measures of dependence between more than two families of random variables. In Section 2.2 below, a special case of Theorem 2.1 is slightly refined.

2.1. *Background and Theorem* 2.1

Suppose $n \geqslant 2$ is a positive integer and for each $k = 1, 2, \ldots, n$, $(\Omega_k, \mathscr{F}_k, P_k)$ is an arbitrary probability space. Throughout Section 2, this n and these probability spaces will be fixed.

For each $k = 1, \ldots, n$ let $\mathscr{S}(\mathscr{F}_k)$ denote the set of all complex-valued $\mathscr{F}_k$-measurable simple functions on Ω_k. If $1 \leqslant k \leqslant n$, then whenever a complex-valued function f is specified to be $\mathscr{F}_k$-measurable, it is understood that f is defined on Ω_k, and that for f the usual p-norms, $1 \leqslant p \leqslant \infty$, are defined with respect to the measure P_k: $\|f\|_p = [\int_{\Omega(k)} |f|^p \, dP_k]^{1/p}$ if $1 \leqslant p < \infty$, and $\|f\|_\infty = P_k\text{-ess sup}\,|f|$. Of course $\|f\|_p \leqslant \|f\|_q$ if $1 \leqslant p \leqslant q \leqslant \infty$ by Hölder's inequality, since P_k is a probability measure. $\mathscr{L}_p(\mathscr{F}_k)$ will denote the set of all complex-valued $\mathscr{F}_k$-measurable functions f on Ω_k such that $\|f\|_p < \infty$.

Suppose $T: \mathscr{S}(\mathscr{F}_1) \times \cdots \times \mathscr{S}(\mathscr{F}_{n-1}) \to \mathscr{L}_1(\mathscr{F}_n)$ is a multilinear operator (or "$(n-1)$-linear" operator). Here "multilinear" or "$(n-1)$-linear" means that for each fixed j, $1 \leqslant j \leqslant n-1$, and each choice of $f_k \in \mathscr{S}(\mathscr{F}_k)$, $k \neq j$, the mapping $T(f_1, \ldots, f_{j-1}, \cdot, f_{j+1}, \ldots, f_{n-1})$ is a linear operator (into $\mathscr{L}_1(\mathscr{F}_n)$). If $n = 2$ then of course T is simply a linear operator.

We shall always make the usual assumption that $T(f_1, \ldots, f_{n-1}) = T(g_1, \ldots, g_{n-1})$ a.e.-P_n if $f_k = g_k$ a.e.-P_k $\forall\, k = 1, \ldots, n-1$, consistent with the usual practice of regarding $\mathscr{L}_p(\mathscr{F}_k)$ as a space of "equivalence classes" of functions.

For each $\mathbf{p} = (p_1, \ldots, p_{n-1}) \in [1, \infty]^{n-1}$ and each $q \in [1, \infty]$, define the following (possibly infinite) norm of T:

$$\|T\|_{\mathbf{p} \to q} := \sup \frac{\|T(f_1, \ldots, f_{n-1})\|_q}{\|f_1\|_{p(1)} \cdot \cdots \cdot \|f_{n-1}\|_{p(n-1)}}, \qquad f_k \in \mathscr{S}(\mathscr{F}_k),\ 1 \leqslant k \leqslant n-1.$$

In this definition, interpret $0/0 = 0$.

It is well known that if $\|T\|_{\mathbf{p} \to q} < \infty$ for given $\mathbf{p}$ and q, then T can be extended uniquely to a bounded multilinear operator from $\mathscr{L}_{p(1)}(\mathscr{F}_1) \times \cdots \times \mathscr{L}_{p(n-1)}(\mathscr{F}_{n-1})$ into $\mathscr{L}_q(\mathscr{F}_n)$, and in this extension the value of the $(\mathbf{p} \to q)$-

norm remains unchanged. But for our purposes it suffices to consider T on just $\mathscr{S}(\mathscr{F}_1) \times \cdots \times \mathscr{S}(\mathscr{F}_{n-1})$.

T is called a "product operator" if $\|T\|_{(\infty,\infty,\ldots,\infty)\to\infty} \leqslant 1$ and $\|T\|_{(\infty,\ldots,\infty,1,\infty,\ldots,\infty)\to 1} \leqslant 1$ for each vector $(\infty,\ldots,\infty,1,\infty,\ldots,\infty)$ (there are $n-1$ of them). This definition is taken from O'Neil [26] and is well known to be motivated partly by the following remark:

Remark 2.1. Let us temporarily abuse notation and denote, for $\mathbf{p} = (p_1,\ldots,p_{n-1}) \in [1,\infty]^{n-1}$, the vector $\mathbf{p}^{-1} := (p_1^{-1},\ldots,p_{n-1}^{-1})$. The multidimensional Riesz–Thorin interpolation theorem (see [1, p. 18, Exercise 13]) says that if $\mathbf{p}_0$, $\mathbf{p}_1$, and $\mathbf{p}$ each $\in [1,\infty]^{n-1}$; q_0, q_1, and q each $\in [1,\infty]$; $0 \leqslant \theta \leqslant 1$; $\mathbf{p}^{-1} = (1-\theta)\,\mathbf{p}_0^{-1} + \theta \mathbf{p}_1^{-1}$; and $q^{-1} = (1-\theta)\,q_0^{-1} + \theta q_1^{-1}$; then $\|T\|_{\mathbf{p}\to q} \leqslant [\|T\|_{\mathbf{p}(0)\to q(0)}]^{1-\theta} \cdot [\|T\|_{\mathbf{p}(1)\to q(1)}]^{\theta}$. By repeated applications, one has $\|T\|_{\mathbf{p}\to q} \leqslant \prod_{m=1}^{M} [\|T\|_{\mathbf{p}(m)\to q(m)}]^{\theta(m)}$ if $\theta_m \geqslant 0$ $\forall\, m$, $\sum_{m=1}^{M} \theta_m = 1$, $\mathbf{p}^{-1} = \sum_{m=1}^{M} \theta_m \mathbf{p}_m^{-1}$, $q^{-1} = \sum_{m=1}^{M} \theta_m q_m^{-1}$, $\mathbf{p}_m \in [1,\infty]^{n-1}$ $\forall\, m$, and $q_m \in [1,\infty]$ $\forall\, m$. In the special case when T is a product operator, this tells us that $\|T\|_{\mathbf{p}\to q} \leqslant 1$ whenever $\mathbf{p} = (p_1,\ldots,p_{n-1}) \in [1,\infty]^{n-1}$, $q \in [1,\infty]$, and $\sum_{k=1}^{n-1} p_k^{-1} = q^{-1}$ (and even when $\sum_{k=1}^{n-1} p_k^{-1} < q^{-1}$, using the fact that P_n is a probability measure).

If we were working with just real-valued functions, then an extra constant factor would have to be incorporated into the above-mentioned interpolation theorem. By working exclusively with complex-valued functions we avoid this extra complication. (See the paragraph following the statement of Theorem 1.3.1 in [1, p. 9].)

One more piece of notation is needed: If $1 \leqslant k \leqslant n$, then whenever a set A is specified to be an element of $\mathscr{F}_k$, $I(A)$ will denote the indicator function of A, defined on Ω_k.

THEOREM 2.1. *Suppose* $\mathbf{p} = (p_1,\ldots,p_{n-1}) \in [1,\infty]^{n-1}$ *and* $1 < q < \infty$ *are such that* $0 \leqslant \sum_{k=1}^{n-1} p_k^{-1} \leqslant q^{-1}$. *Then there exists a constant* $C = C(\mathbf{p}; q)$ *which is a function only of* $\mathbf{p}$ *and* q, *such that the following statement holds*:

Suppose $T\colon \mathscr{S}(\mathscr{F}_1) \times \cdots \times \mathscr{S}(\mathscr{F}_{n-1}) \to \mathscr{L}_{\infty}^{c}(\mathscr{F}_n)$ *is an* $(n-1)$*-linear product operator; suppose that* $\forall\, k$, $1 \leqslant k \leqslant n-1$, *either* $\mathscr{G}_k := \mathscr{S}(\mathscr{F}_k)$ *or* $\mathscr{G}_k := \{I(A)\colon A \in \mathscr{F}_k\}$; *suppose that* $0 < \varepsilon \leqslant 1$ *and that* $\forall\, t > 0$, $\forall\, (f_1,\ldots,f_{n-1}) \in \mathscr{G}_1 \times \cdots \times \mathscr{G}_{n-1}$ *one has* $P_n(|T(f_1,\ldots,f_{n-1})| > t) \leqslant [(\varepsilon/t) \cdot \prod_{k=1}^{n-1} \|f_k\|_{p(k)}]^q$; *then* $\forall\, (f_1,\ldots,f_{n-1}) \in \mathscr{G}_1 \times \cdots \times \mathscr{G}_{n-1}$ *one has* $\|T(f_1,\ldots,f_{n-1})\|_q \leqslant [C \cdot \varepsilon \cdot (1-\log\varepsilon)^{1/q}] \cdot \prod_{k=1}^{n-1} \|f_k\|_{p(k)}$.

It is emphasized that C does not depend at all on the particular probability spaces $(\Omega_k, \mathscr{F}_k, P_k)$ being used; we shall sometimes write it as $C(p_1,\ldots,p_{n-1}; q)$ when we wish to mention the components of $\mathbf{p}$ explicitly. When $\mathscr{G}_k = \mathscr{S}(\mathscr{F}_k)$ $\forall\, k$, the last inequality simply says $\|T\|_{\mathbf{p}\to q} \leqslant C \cdot \varepsilon \cdot (1-\log\varepsilon)^{1/q}$. The use of $\mathscr{G}_k$ at all (i.e., allowing the option $\mathscr{G}_k = \{I(A),\ A \in \mathscr{F}_k\}$ for some or all k's) is only a slight extra complication, and will facilitate the use of Theorem 2.1 later on. Of course $1-\log\varepsilon \geqslant 1$ since $\varepsilon \leqslant 1$. When A. Torchinsky first saw the statement of Theorem 4.3 in Section 4.4 (which had been proved earlier), he conjectured that there was a connection between such results and the work on BMO functions in harmonic analysis. Indeed, the proof of Theorem 2.1 will make use of techniques well known in the study of BMO functions. Similar results with $q=1$ or ∞ will not be considered here. In the special case where $n=2$ and $p_1=q=2$, Theorem 2.1 is sharp up to a constant factor; see Remark 4.5 in Section 4.4.

Remark 2.2. In Theorem 2.1, if $\varepsilon=1$ then the last inequality obviously becomes trivial with C replaced by 1, since T is a product operator.

Proof of Theorem 2.1. Assume $\mathbf{p}$ and q are as in the statement of Theorem 2.1. We shall break the argument into three cases:

Case I: $0<\sum_{k=1}^{n-1} p_k^{-1}=q^{-1}$ and $p_k<\infty\ \forall\, k$.

Case II: $0<\sum_{k=1}^{n-1} p_k^{-1}=q^{-1}$ and $p_k=\infty$ for some k.

Case III: $0\leqslant\sum_{k=1}^{n-1} p_k^{-1}<q^{-1}$.

Case I is the critical one. Let us take it for granted for a moment and quickly get Cases II and III out of the way with simple arguments.

Proof for Case II. Permuting indices if necessary, we may assume without loss of generality that for some m, $1\leqslant m\leqslant n-2$, one has $p_k=\infty$ $\forall\, k\leqslant m$ and $1<p_k<\infty\ \forall\, k\geqslant m+1$.

Now suppose that T, $\mathscr{G}_1,\ldots,\mathscr{G}_{n-1}$, and ε fulfill the assumptions in the statement of Theorem 2.1.

Let the functions $f_k\in\mathscr{G}_k$, $1\leqslant k\leqslant m$, be arbitrary but fixed, and define the $(n-1-m)$-linear operator $T'\colon \mathscr{S}(\mathscr{F}_{m+1})\times\cdots\times\mathscr{S}(\mathscr{F}_{n-1})\to\mathscr{L}_\infty(\mathscr{F}_n)$ by $T'(g_1,\ldots,g_{n-1-m}):=(\prod_{k=1}^m\|f_k\|_\infty)^{-1}\,T(f_1,\ldots,f_m,g_1,\ldots,g_{n-1-m})$ ($T':=0$ if $f_k=0$ a.e.-P_k for some $k\leqslant m$). Note that T' is a product operator. A simple argument shows that $\forall\, t>0$, $\forall\,(g_1,\ldots,g_{n-1-m})\in\mathscr{G}_{m+1}\times\cdots\times\mathscr{G}_{n-1}$, one has $P_n(|T'(g_1,\ldots,g_{n-1-m})|>t)\leqslant[(\varepsilon/t)\cdot\prod_{k=1}^{n-1-m}\|g_k\|_{p(k+m)}]^q$. Hence by Theorem 2.1 for Case I, $\forall\,(g_1,\ldots,g_{n-1-m})\in\mathscr{G}_{m+1}\times\cdots\times\mathscr{G}_{n-1}$, one has

$$\begin{aligned}\|T(f_1,\ldots,f_m,g_1,\ldots,g_{n-1-m})\|_q&=\left(\prod_{k=1}^m\|f_k\|_\infty\right)\cdot\|T'(g_1,\ldots,g_{n-1-m})\|_q\\&\leqslant[C(p_{m+1},\ldots,p_{n-1};q)\cdot\varepsilon\cdot(1-\log\varepsilon)^{1/q}]\\&\quad\cdot\left(\prod_{k=1}^m\|f_k\|_\infty\right)\cdot\left(\prod_{k=1}^{n-1-m}\|g_k\|_{p_{k+m}}\right).\end{aligned}$$

Since $f_k\in\mathscr{G}_k$, $1\leqslant k\leqslant m$, were also arbitrary, Theorem 2.1 holds for Case II, with $C(\infty,\ldots,\infty,p_{m+1},\ldots,p_{n-1};q)=C(p_{m+1},\ldots,p_{n-1};q)$.

Proof for Case III. Suppose that T, $\mathscr{G}_1,\ldots,\mathscr{G}_{n-1}$, and ε fulfill the assumptions in the statement of Theorem 2.1. Define p_0, $1 < p_0 < \infty$, such that $\sum_{k=0}^{n-1} p_k^{-1} = q^{-1}$. Let $(\Omega_0, \mathscr{F}_0, P_0)$ be a trivial probability space with Ω_0 consisting of just one element. Define the n-linear operator $T': \mathscr{S}(\mathscr{F}_0) \times \cdots \times \mathscr{S}(\mathscr{F}_{n-1}) \to \mathscr{L}_\infty(\mathscr{F}_n)$ by $T'(f_0,\ldots,f_{n-1}) := f_0 \cdot T(f_1,\ldots,f_{n-1})$ (f_0 is of course just a scalar). Taking Cases I and II for granted and mimicking the argument for Case II (more or less), one can now derive Theorem 2.1 for Case III as a consequence of Theorem 2.1 for Case I or II (whichever is applicable), with $C(p_1,\ldots,p_{n-1}; q) = C(p_0,\ldots,p_{n-1}; q)$.

Now we only need to establish Theorem 2.1 for Case I.

Proof for Case I. Let $\mathbf{p} = (p_1,\ldots,p_{n-1}) \in [1, \infty]^{n-1}$ and q be fixed with $0 < \sum_{k=1}^{n-1} p_k^{-1} = q^{-1} < 1$ and $p_k < \infty\ \forall\, k$.

First let us define the quantity C, along with some other parameters that will be needed later:

$$\begin{aligned}
\theta &:= q-1,\\
\lambda &:= \min_{1 \leqslant k \leqslant n-1} (2p_k)^{-1},\\
w &:= q/(\lambda\theta), \qquad\qquad (2.1)\\
C = C(\mathbf{p}, q) &:= [3^{(q+1)(n-1)} q \cdot 2w \cdot \max\{1, \theta^{-1}\}]^{1/q},\\
\alpha_k &:= p_k/q \qquad \forall\, k = 1,\ldots, n-1.
\end{aligned}$$

Every parameter defined here is clearly positive and depends only on $\mathbf{p}$ and q. Note the following trivial facts:

$$\begin{gathered}
w > 1,\\
p_k - \theta\alpha_k = p_k/q \geqslant 1 \qquad \forall\, k = 1,\ldots, n-1, \qquad\qquad (2.2)\\
\sum_{k=1}^{n-1} 1/\alpha_k = 1 = \sum_{k=1}^{n-1} q/p_k.
\end{gathered}$$

Now suppose that T, $\mathscr{G}_1,\ldots,\mathscr{G}_{n-1}$, and ε satisfy the assumptions in the statement of Theorem 2.1 (with respect to $\mathbf{p}$ and q).

Let $f_k \in \mathscr{G}_k$, $1 \leqslant k \leqslant n-1$, be arbitrary functions such that $\|f_k\|_\infty > 0\ \forall\, k$. Define the functions g_k, $1 \leqslant k \leqslant n-1$, by $g_k := f_k/\|f_k\|_{p(k)}$. In what follows, we shall work with g_k instead of f_k because of the convenient property $\|g_k\|_{p(k)} = 1\ \forall\, k$. To prove Theorem 2.1 it suffices to prove

$$\|T(g_1,\ldots,g_{n-1})\|_q \leqslant C \cdot \varepsilon \cdot (1 - \log \varepsilon)^{1/q}. \qquad (2.3)$$

For each k, $1 \leqslant k \leqslant n-1$, and each $t > 0$, define the following three functions on Ω_k:

$$\begin{aligned} g_{k,1}^{(t)} &:= g_k \cdot I(\varepsilon^{-w} \cdot |g_k|^{\alpha(k)} < t), \\ g_{k,2}^{(t)} &:= g_k \cdot I(\varepsilon^{w} \cdot |g_k|^{\alpha(k)} < t \leqslant \varepsilon^{-w} \cdot |g_k|^{\alpha(k)}), \\ g_{k,3}^{(t)} &:= g_k \cdot I(t \leqslant \varepsilon^{w} \cdot |g_k|^{\alpha(k)}). \end{aligned} \tag{2.4}$$

(See (2.1), and keep in mind that $\varepsilon^w \leqslant \varepsilon^{-w}$ since $\varepsilon \leqslant 1$.) Then for each k and each $t > 0$, $g_k = g_{k,1}^{(t)} + g_{k,2}^{(t)} + g_{k,3}^{(t)}$.

Let $S = \{(i_1, ..., i_{n-1}): i_k \in \{1, 2, 3\}\ \forall\, k\}$. (Thus S has 3^{n-1} elements.) Then for each $t > 0$, $T(g_1, ..., g_{n-1}) = \sum_{(i(1), ..., i(n-1)) \in S} T(g_{1,i(1)}^{(t)}, ..., g_{n-1,i(n-1)}^{(t)})$. Hence for each $t > 0$ we have the following inclusion of events (sets in $\mathscr{F}_n$):

$$\{|T(g_1, ..., g_{n-1})| > 3^{n-1} t\} \subset \bigcup_S \{|T(g_{1,i(1)}^{(t)}, ..., g_{n-1,i(n-1)}^{(t)})| > t\}.$$

In what follows, if $1 \leqslant k \leqslant n$ and $f \in \mathscr{L}_1(\mathscr{F}_k)$ then the notation $E_k f$ means $\int_{\Omega(k)} f\, dP_k$.

Now by a well known identity and a simple substitution,

$$\begin{aligned} E_n |T(g_1, ..., g_{n-1})|^q &= q \cdot \int_0^\infty t^{q-1} P_n(|T(g_1, ..., g_{n-1})| > t)\, dt \\ &= 3^{q(n-1)} q \cdot \int_0^\infty t^{q-1} P_n(|T(g_1, ..., g_{n-1})| > 3^{n-1} t)\, dt \\ &\leqslant 3^{q(n-1)} q \cdot \sum_S I_{i(1), ..., i(n-1)} \end{aligned} \tag{2.5}$$

where $\forall\, (i_1, ..., i_{n-1}) \in S$,

$$I_{i(1), ..., i(n-1)} := \int_0^\infty t^{q-1} P_n(|T(g_{1,i(1)}^{(t)}, ..., g_{n-1,i(n-1)}^{(t)})| > t)\, dt.$$

In Lemmas 1 and 2 below, we shall derive upper bounds on these numbers $I_{i(1), ..., i(n-1)}$.

LEMMA 1. *For each* $(i_1, ..., i_{n-1}) \in S - \{(2, 2, ..., 2)\}$ *one has* $I_{i(1), ..., i(n-1)} \leqslant 2w \cdot (\max\{1, \theta^{-1}\}) \cdot \varepsilon^q (1 - \log \varepsilon)$.

Proof. Let $(i_1, ..., i_{n-1}) \in S$ be fixed ($\neq (2, 2, ..., 2)$). Define r_k, $1 \leqslant k \leqslant n-1$, by

$$\begin{aligned} r_k &= p_k + \theta \alpha_k && \text{if } i_k = 1 \\ &= p_k && \text{if } i_k = 2 \\ &= p_k - \theta \alpha_k && \text{if } i_k = 3. \end{aligned}$$

By (2.2), $\sum_{k=1}^{n-1} r_k^{-1} \leqslant \sum_{k=1}^{n-1}(p_k - \theta\alpha_k)^{-1} = 1$. Define $\gamma \geqslant 1$ by $\gamma^{-1} = \sum_{k=1}^{n-1} r_k^{-1}$.

By the Markov inequality, $\forall\, t > 0$,

$$\begin{aligned} P_n(|T(g_{1,i(1)}^{(t)}, \dots, g_{n-1,i(n-1)}^{(t)})| > t) &\leqslant t^{-\gamma} E_n\, |T(g_{1,i(1)}^{(t)}, \dots, g_{n-1,i(n-1)}^{(t)})|^{\gamma} \\ &\leqslant t^{-\gamma} \cdot \prod_{k=1}^{n-1} \|g_{k,i(k)}^{(t)}\|_{r(k)}^{\gamma} \end{aligned} \tag{2.6}$$

since T is a product operator. Hence

$$I_{i(1),\dots,i(n-1)} \leqslant \int_0^\infty t^{q-1-\gamma} \cdot \left(\prod_{k=1}^{n-1} \|g_{k,i(k)}^{(t)}\|_{r(k)}^{\gamma} \right) dt. \tag{2.7}$$

Define u_k, $1 \leqslant k \leqslant n-1$, by

$$\begin{aligned} u_k &= 1 + \theta && \text{if } i_k = 1 \\ &= 1 && \text{if } i_k = 2 \\ &= 1 - \theta && \text{if } i_k = 3. \end{aligned}$$

Then $\forall\, k = 1, \dots, n-1$, $r_k + (-u_k + 1)\,\alpha_k = p_k$ and hence by (2.1), $-u_k = -1 + (p_k/\alpha_k) - (r_k/\alpha_k) = -1 + q - (r_k/\alpha_k)$. Hence by (2.2),

$$\begin{aligned} -\sum_{k=1}^{n-1} u_k \gamma / r_k &= (q-1) \sum_{k=1}^{n-1} \gamma / r_k - \gamma \sum_{k=1}^{n-1} 1/\alpha_k \\ &= (q-1) \cdot 1 - \gamma \cdot 1 = q - 1 - \gamma. \end{aligned}$$

Hence by the (multidimensional) Hölder inequality, Fubini's theorem, and (2.7),

$$\begin{aligned} I_{i(1),\dots,i(n-1)} &\leqslant \int_0^\infty \prod_{k=1}^{n-1} [t^{-u(k)} \cdot \|g_{k,i(k)}^{(t)}\|_{r(k)}^{r(k)}]^{\gamma/r(k)}\, dt \\ &\leqslant \prod_{k=1}^{n-1} \left[\int_0^\infty t^{-u(k)} E_k\, |g_{k,i(k)}^{(t)}|^{r(k)}\, dt \right]^{\gamma/r(k)} \\ &= \prod_{k=1}^{n-1} \left[E_k \int_0^\infty t^{-u(k)}\, |g_{k,i(k)}^{(t)}|^{r(k)}\, dt \right]^{\gamma/r(k)}. \end{aligned} \tag{2.8}$$

For any k, $1 \leqslant k \leqslant n-1$, such that $i_k = 1$ one has

$$
\begin{aligned}
E_k \int_0^\infty t^{-u(k)} \, |g_{k,i(k)}^{(t)}|^{r(k)} \, dt &= E_k \int_{\varepsilon^{-w}|g_k|^{\alpha(k)}}^\infty t^{-u(k)} \, |g_k|^{r(k)} \, dt \\
&= E_k \left[|g_k|^{r(k)} \int_{\varepsilon^{-w}|g_k|^{\alpha(k)}}^\infty t^{-1-\theta} \, dt \right] \\
&= E_k [|g_k|^{r(k)} \cdot \theta^{-1} \cdot (\varepsilon^{-w} \, |g_k|^{\alpha(k)})^{-\theta}] \\
&= E_k [|g_k|^{p(k)} \cdot \theta^{-1} \varepsilon^{\theta w}] = \theta^{-1} \varepsilon^{\theta w}. \qquad (2.9)
\end{aligned}
$$

Similarly, for any k such that $i_k = 2$ one has

$$
\begin{aligned}
&E_k \int_0^\infty t^{-u(k)} \, |g_{k,i(k)}^{(t)}|^{r(k)} \, dt \\
&\quad = E_k \left[|g_k|^{r(k)} \cdot \int_{\varepsilon^{w}|g_k|^{\alpha(k)}}^{\varepsilon^{-w}|g_k|^{\alpha(k)}} t^{-1} \, dt \right] = -2w \log \varepsilon \qquad (2.10)
\end{aligned}
$$

and for any k such that $i_k = 3$ one has

$$
\begin{aligned}
&E_k \int_0^\infty t^{-u(k)} \, |g_{k,i(k)}^{(t)}|^{r(k)} \, dt \\
&\quad = E_k \left[|g_k|^{r(k)} \cdot \int_0^{\varepsilon^{w}|g_k|^{\alpha(k)}} t^{-1+\theta} \, dt \right] \\
&\quad = \theta^{-1} \varepsilon^{\theta w}. \qquad (2.11)
\end{aligned}
$$

By (2.8)–(2.11), $I_{i(1),\ldots,i(n-1)} \leqslant (\theta^{-1}\varepsilon^{\theta w})^{a(1)+a(3)} \cdot (-2w \log \varepsilon)^{a(2)}$, where for $m = 1, 2, 3$, $a(m) := \sum_{\{k: i(k) = m\}} \gamma / r_k$. Now $a(1) + a(2) + a(3) = 1$; hence $(\theta^{-1})^{a(1)+a(3)} \leqslant \max\{1, \theta^{-1}\}$; hence $I_{i(1),\ldots,i(n-1)} \leqslant (\max\{1, \theta^{-1}\}) \cdot \varepsilon^{\theta w[a(1)+a(3)]} \cdot (-2w \log \varepsilon)^{a(2)}$. Now $i(k) \neq 2$ for some k by the hypothesis of Lemma 1, and for such a k we have by (2.1), $\gamma / r_k \geqslant 1/r_k \geqslant 1/(p_k + \theta\alpha_k) \geqslant \lambda$, and hence $\theta w[a(1) + a(3)] \geqslant \theta w \lambda = q$. Since $\varepsilon \leqslant 1$, the inequality in Lemma 1 holds.

LEMMA 2. $I_{(2,2,\ldots,2)} \leqslant 2w \cdot \varepsilon^q (1 - \log \varepsilon)$.

Proof. For each k and t, one has $g_{k,2}^{(t)} = c \cdot h$ for some $h \in \mathscr{G}_k$ and some positive number c, regardless of whether $\mathscr{G}_k = \mathscr{S}(\mathscr{F}_k)$ or $\mathscr{G}_k = \{I(A)\colon A \in \mathscr{F}_k\}$. From the assumptions in Theorem 2.1 and a trivial argument one has that $\forall\, t > 0$, $P_n(|T(g_{1,2}^{(t)}, g_{2,2}^{(t)}, \ldots, g_{n-1,2}^{(t)})| > t) \leqslant [(\varepsilon/t) \cdot \prod_{k=1}^{n-1} \|g_{k,2}^{(t)}\|_{p(k)}]^q$.

Since $\sum_{k=1}^{n-1} q/p_k = 1$ (see (2.2)), we get

$$
\begin{aligned}
I_{(2,2,\ldots,2)} &\leqslant \int_0^\infty t^{q-1}\left[(\varepsilon/t)\cdot\prod_{k=1}^{n-1}\|g_{k,2}^{(t)}\|_{p(k)}\right]^q dt \\
&\leqslant \varepsilon^q \cdot \prod_{k=1}^{n-1}\left[\int_0^\infty t^{-1}E_k\,|g_{k,2}^{(t)}|^{p(k)}\,dt\right]^{q/p(k)} \\
&= \varepsilon^q \cdot \prod_{k=1}^{n-1}\left[E_k \int_{\varepsilon^w |g_k|^{\alpha(k)}}^{\varepsilon^{-w}|g_k|^{\alpha(k)}} t^{-1}\,|g_k|^{p(k)}\,dt\right]^{q/p(k)} \\
&= \varepsilon^q \cdot \prod_{k=1}^{n-1}\left[E_k\,|g_k|^{p(k)}(-2w\log\varepsilon)\right]^{q/p(k)} \\
&= \varepsilon^q \cdot (-2w\log\varepsilon). \qquad (2.12)
\end{aligned}
$$

Thus Lemma 2 holds.

Now by (2.5) and Lemmas 1 and 2 we have

$$E_n\,|T(g_1,\ldots,g_{n-1})|^q \leqslant 3^{q(n-1)}q\cdot 2w[3^{n-1}\max\{1,\theta^{-1}\}]\cdot\varepsilon^q(1-\log\varepsilon)$$

and hence (2.3) holds by (2.1). This completes the proof of Theorem 2.1.

2.2. *A Refinement*

In the proof of Theorem 2.1, with a little more flexibility in the arguments one can produce a lower value for C than the value given in (2.1). This is particularly true for small values of n. Here we shall illustrate this for just one special case:

THEOREM 2.2. *In the special case where $n=2$ and $1<p_1=q<\infty$, Theorem 2.1 holds with $C=C(q;q)=3\cdot[q^2/(q-1)]^{1/q}$.*

Proof. We shall carry out the argument of Theorem 2.1 with a few minor modifications. We shall ignore (2.1) and (2.2). In place of (2.3) we shall prove

$$\|T(g_1)\|_q \leqslant 3\cdot[q^2/(q-1)]^{1/q}\cdot\varepsilon\cdot(1-\log\varepsilon)^{1/q}. \qquad (2.3')$$

Theorem 2.2 will then follow (because we shall have $\|g_1\|_q=1$).

As was noted in Remark 2.2, we can dismiss the case $\varepsilon=1$ and assume $\varepsilon<1$. Fix p very large, $q<p<\infty$. Define the constants $A := \varepsilon^{q/(q-1)}$ and $D := (p-q)^{1/(q-p)}\cdot\varepsilon^{q/(q-p)}$. Replace (2.4) by

$$
\begin{aligned}
g_{1,1}^{(t)} &:= g_1\cdot I(D\,|g_1|<t),\\
g_{1,2}^{(t)} &:= g_1\cdot I(A\,|g_1|<t\leqslant D\,|g_1|), \qquad (2.4')\\
g_{1,3}^{(t)} &:= g_1\cdot I(t\leqslant A\,|g_1|).
\end{aligned}
$$

We shall assume $D > A$. We can insure this simply by taking p sufficiently large, since $A < 1$ (because $\varepsilon < 1$) and $\mathrm{Lim}_{p \to \infty} D = 1$. (In fact we shall let $p \to \infty$ later on.)

Now (2.5) becomes

$$E_2 \, |T(g_1)|^q \leqslant 3^q q \cdot (I_1 + I_2 + I_3). \tag{2.5'}$$

Arguing essentially as in (2.6) and (2.7) we have that for each γ, $1 \leqslant \gamma < \infty$, and each $i(1) \in \{1, 2, 3\}$,

$$I_{i(1)} \leqslant \int_0^\infty t^{q-1-\gamma} E_1 \, |g_{1,i(1)}^{(t)}|^\gamma \, dt. \tag{2.7'}$$

Starting with (2.7′) with $\gamma = p$ and imitating the argument of (2.9), we get $I_1 \leqslant D^{q-p}/(p-q)$. Setting $\gamma = 1$ in (2.7′) and imitating (2.11), we get $I_3 \leqslant A^{q-1}/(q-1)$. Also, by imitating (2.12) we get $I_2 \leqslant \varepsilon^q \log(D/A)$. Plugging these into (2.5′) we get $E_2 \, |T(g_1)|^q \leqslant 3^q q[(q/(q-1)) \, \varepsilon^q + \varepsilon^q \log(D/A)]$. Finally, using $\mathrm{Lim}_{p \to \infty} D = 1$ and elementary arithmetic, we obtain (2.3′). This completes the proof of Theorem 2.2.

III. Multilinear Forms

This is a continuation of Section 2. As in Section 2 we fix an arbitrary integer $n \geqslant 2$ and arbitrary probability spaces $(\Omega_k, \mathscr{F}_k, P_k)$. The other notations and definitions in Section 2 are also carried over.

Suppose $B\colon \mathscr{S}(\mathscr{F}_1) \times \cdots \times \mathscr{S}(\mathscr{F}_n) \to \mathbb{C}$ is a multilinear form (or "n-linear" form), where $\mathbb{C}$ denotes the field of complex numbers. This terminology means of course that for each fixed j, $1 \leqslant j \leqslant n$, and each fixed choice of functions $f_k \in \mathscr{S}(\mathscr{F}_k)$, $k \neq j$, the mapping $B(f_1, ..., f_{j-1}, \cdot, f_{j+1}, ..., f_n)$ is a (complex) linear functional on $\mathscr{S}(\mathscr{F}_j)$. As usual, we assume that $B(f_1, ..., f_n) = B(g_1, ..., g_n)$ if $f_k = g_k$ a.e.-P_k $\forall\, k$.

For our particular discussion of measures of dependence in Sections 1 and 4, and especially for the measures of dependence between three or more families of random variables as discussed in Section 4.2, it seems more natural to work with multilinear forms than with the multilinear operators studied in Section 2. Section 3 is devoted to comparisons between norms for n-linear forms; these norms will be closely related to the measures of dependence discussed in Sections 1 and 4. Here we shall present six theorems. The first five are all either trivial or well known, and the sixth will be (ultimately) a consequence of Theorem 2.1. All six theorems will be useful in Section 4.

For each n-linear form B on $\mathscr{S}(\mathscr{F}_1)\times\cdots\times\mathscr{S}(\mathscr{F}_n)$ and each $\mathbf{p}=(p_1,\dots,p_n)\in[1,\infty]^n$, define the following (possibly infinite) quantities:

$$\|B\|_{\mathbf{p}}:=\sup\frac{|B(f_1,\dots,f_n)|}{\prod_{k=1}^n\|f_k\|_{p(k)}},\qquad f_k\in\mathscr{S}(\mathscr{F}_k),\ k=1,\dots,n,$$

$$d_{\mathbf{p}}(B):=\sup\frac{|B(I(A_1),\dots,I(A_n))|}{\prod_{k=1}^n[P_k(A_k)]^{1/p(k)}},\qquad A_k\in\mathscr{F}_k,\ k=1,\dots,n.$$

(Again interpret $0/0=0$.) Of course $d_{\mathbf{p}}(B)\leqslant\|B\|_{\mathbf{p}}$ for any given $\mathbf{p}$, since $\|I(A_k)\|_{p(k)}=[P_k(A_k)]^{1/p(k)}$. Note the simple equality $d_{(\infty,\infty,\dots,\infty)}(B)=\sup\{|B(I(A_1),\dots,I(A_n))|: A_k\in\mathscr{F}_k\ \forall\, k\}$. $\|B\|_{\mathbf{p}}$ is the usual $\mathbf{p}$-norm of B. In the case $n=2$, Stein and Weiss [37] studied the conditions $d_{\mathbf{p}}(B)<\infty$ and similar conditions on linear operators; following their terminology one might refer to $d_{\mathbf{p}}(B)$ as a "restricted" norm.

We need some notation for certain special vectors in $[1,\infty]^n$: $\boldsymbol{\infty}:=(\infty,\infty,\dots,\infty)$; and $\forall\, k=1,\dots,n$, $\mathbf{i}_k:=(\infty,\dots,\infty,1,\infty,\dots,\infty)$, where the 1 is the kth coordinate.

THEOREM 3.1. *Suppose B is an n-linear form, $\mathbf{p}=(p_1,\dots,p_n)\in[1,\infty]^n$, and $s:=\sum_{k=1}^n p_k^{-1}\leqslant 1$. Then $d_{\boldsymbol{\infty}}(B)\leqslant d_{\mathbf{p}}(B)\leqslant[d_{\boldsymbol{\infty}}(B)]^{1-s}\cdot\prod_{k=1}^n[d_{\mathbf{i}(k)}(B)]^{1/p(k)}$.*

THEOREM 3.2. *Suppose B is an n-linear form. Suppose $\mathbf{p}=(p_1,\dots,p_n)\in[1,\infty]^n$ and $\mathbf{q}=(q_1,\dots,q_n)\in[1,\infty]^n$ such that $\sum_{k=1}^n p_k^{-1}=\sum_{k=1}^n q_k^{-1}=1$ and $\{k: p_k=\infty\}\subset\{k: q_k=\infty\}$. Then $d_{\mathbf{p}}(B)\leqslant[d_{\mathbf{q}}(B)]^{\theta}\cdot\prod_{k=1}^n[d_{\mathbf{i}(k)}(B)]^{\alpha(k)}$, where $\theta:=\min_{\{k:q(k)<\infty\}}(q_k/p_k)$ and $\forall\, k$, $\alpha(k):=p_k^{-1}-\theta q_k^{-1}$.*

In Theorem 3.2 note that $0<\theta\leqslant 1$, $\alpha(k)\geqslant 0\ \forall\, k$, and $\theta+\sum_{k=1}^n\alpha(k)=1$. The same comment will apply to Theorem 3.4 below. To prove Theorems 3.1 and 3.2 one can simply note that, for fixed $A_1\in\mathscr{F}_1,\dots,A_n\in\mathscr{F}_n$, defining $b:=|B(I(A_1),\dots,I(A_n))|$, in the context of Theorem 3.1 one has

$$b\leqslant\frac{b}{\prod_{k=1}^n[P_k(A_k)]^{1/p(k)}}=b^{1-s}\cdot\prod_{k=1}^n\left[\frac{b}{P_k(A_k)}\right]^{1/p(k)}$$

and in the context of Theorem 3.2 one has

$$\frac{b}{\prod_{k=1}^n[P_k(A_k)]^{1/p(k)}}=\left[\frac{b}{\prod_{k=1}^n[P_k(A_k)]^{1/q(k)}}\right]^{\theta}\cdot\prod_{k=1}^n\left[\frac{b}{P_k(A_k)}\right]^{\alpha(k)}.$$

Theorems 3.1 and 3.2 follow easily. (In the theorems in this section, we shall not labor over such trivial cases as, say, when $d_{\mathbf{p}}(B)=\infty$ or 0.)

THEOREM 3.3. *Suppose B, $\mathbf{p}$, and s are as in Theorem 3.1. Then* $\|B\|_\infty \leqslant \|B\|_\mathbf{p} \leqslant [\|B\|_\infty]^{1-s} \cdot \prod_{k=1}^n [\|B\|_{\mathbf{i}(k)}]^{1/p(k)}$.

THEOREM 3.4. *Suppose B, $\mathbf{p}$, $\mathbf{q}$, θ, and $\alpha(k)$, $1 \leqslant k \leqslant n$, are as in Theorem 3.2. Then* $\|B\|_\mathbf{p} \leqslant [\|B\|_\mathbf{q}]^\theta \cdot \prod_{k=1}^n [\|B\|_{\mathbf{i}(k)}]^{\alpha(k)}$.

Theorems 3.3 and 3.4 are just simple consequences of the multidimensional Riesz–Thorin interpolation theorem. (Apply the argument in Remark 2.1, but in using [1, p. 18, Exercise 13] think of B as a linear operator into $\mathscr{L}_1(\Omega_{n+1}, \mathscr{F}_{n+1}, P_{n+1})$, where Ω_{n+1} is trivial, consisting of just one element.)

THEOREM 3.5. *Suppose B is an n-linear form. Then*

(i) $\|B\|_\infty \leqslant 6^n \cdot d_\infty(B)$;

(ii) $\forall\, k = 1,..., n$, $\|B\|_{\mathbf{i}(k)} \leqslant 6^{n-1} \cdot d_{\mathbf{i}(k)}(B)$; *and*

(iii) $\|B\|_{(1,1,...,1)} = d_{(1,1,...,1)}(B)$.

In different guises this theorem has been used frequently in probability theory (e.g., in [2, 12, 18, 20, 21, 33, 38]). Its proof will be postponed until after the statement of Lemma 3.7 below. The reader seeking the sharpest possible constants to replace 6^n and 6^{n-1} (which are not sharp) might find [12, p. 528, Lemma 5.3] and [21, p. 121, Lemma 5] to be valuable.

Before stating Theorem 3.6 we need another definition: An n-linear form B is a "product form" if $\|B\|_{\mathbf{i}(k)} \leqslant 1$ $\forall\, k = 1,..., n$. Of course by Theorem 3.3 (and analogous to Remark 2.1), $\|B\|_\mathbf{p} \leqslant 1$ whenever B is a product form and $\sum_{k=1}^n p_k^{-1} \leqslant 1$.

THEOREM 3.6. *Suppose $\mathbf{p} = (p_1,..., p_n) \in [1, \infty]^n$ such that $\sum_{k=1}^n p_k^{-1} \leqslant 1$. Define the number $c = c(\mathbf{p}) :=$ (cardinality of $\{k\colon p_k < \infty\}) - \sum_{k=1}^n p_k^{-1}$. Then there exists a constant $C = C(\mathbf{p})$ which is a function only of $\mathbf{p}$, such that the following statement holds:*

If B is an n-linear product form then $\|B\|_\mathbf{p} \leqslant C \cdot d_\mathbf{p}(B) \cdot [1 - \log d_\mathbf{p}(B)]^c$.

It is emphasized that the constant C, like c, does not depend at all on the particular probability spaces $(\Omega_k, \mathscr{F}_k, P_k)$, $1 \leqslant k \leqslant n$, being used. The proof of Theorem 3.6 will be postponed until after the statement of Lemma 3.7 below.

Remark 3.1. Theorem 3.6 is closely related to the work of Stein and Weiss [37]. In the context of Theorem 3.6 for the case $n = 2$, if $\mathbf{p} = (p_1, p_2)$, where $1 < p_1$, $p_2 < \infty$ and $p_1^{-1} + p_2^{-1} \leqslant 1$, the slightly weaker inequality $\|B\|_\mathbf{p} \leqslant C \cdot [d_\mathbf{p}(B)]^{1-\varepsilon}$, where $\varepsilon > 0$ can be fixed arbitrarily small and C depends only on ε and $\mathbf{p}$, can be established (for product forms) by carrying out two applications of [37, Theorem VII], with the parameters chosen carefully depending on $\mathbf{p}$ and ε, followed by an application of the Riesz interpolation (convexity) theorem. In order to do this, one first has to incorporate explicitly into [37, Theorem VII] a bound which is based partly on the arguments in [37] and partly on a bound in the Marcinkiewicz interpolation theorem. In any case this approach is quicker than the argument used in [3]. Theorem 3.6 is proved by combining (in the proof of Lemma 3.7 below) Theorem 2.1 and a key idea in [37].

Remark 3.2. In our applications of the theorems in this section to measures of dependence in Section 4, we shall be primarily interested in product forms (or n-linear forms which differ from a product form by only a constant factor). If $B_1, B_2, \ldots$ is a sequence of (n-linear) product forms and $\mathbf{p}=(p_1, \ldots, p_n) \in [1, \infty]^n$ with $\sum_{k=1}^n p_k^{-1} < 1$, then $\operatorname{Lim}_{j\to\infty} d_{\mathbf{p}}(B_j)=0$ if and only if $\operatorname{Lim}_{j\to\infty} d_{\infty}(B_j)=0$ by Theorem 3.1, and thus for product forms the norms $d_{\mathbf{p}}$ and d_∞ can in a certain sense be regarded as "equivalent." In this sense, for product forms, by Theorems 3.1–3.6, the norms $d_{\mathbf{p}}(\cdot)$, $\|\cdot\|_{\mathbf{p}}$, $d_{\mathbf{q}}(\cdot)$, and $\|\cdot\|_{\mathbf{q}}$ are equivalent, where $\mathbf{p}=(p_1, \ldots, p_n) \in [1, \infty]^n$ and $\mathbf{q}=(q_1, \ldots, q_n) \in [1, \infty]^n$, if either (i) $\sum_{k=1}^n p_k^{-1} < 1$ and $\sum_{k=1}^n q_k^{-1} < 1$, or else (ii) $\sum_{k=1}^n p_k^{-1} = \sum_{k=1}^n q_k^{-1} = 1$ and $\{k\colon p_k=\infty\} = \{k\colon q_k=\infty\}$.

The proofs of Theorems 3.5 and 3.6 will be based on the following cumbersome technical lemma:

LEMMA 3.7. *Suppose B is an n-linear form and $\mathbf{p}=(p_1, \ldots, p_n) \in [1, \infty]^n$. Suppose $1 \leqslant j \leqslant n$; and suppose that for each k, $1 \leqslant k \leqslant n$, $k \neq j$, either $\mathscr{G}_k := \mathscr{S}(\mathscr{F}_k)$ or $\mathscr{G}_k := \{I(A)\colon A \in \mathscr{F}_k\}$. Define the quantities $\Delta^{(1)}$ and $\Delta^{(2)}$ by*

$$\Delta^{(1)} := \sup \frac{|B(f_1, \ldots, f_n)|}{\|f_1\|_{p(1)} \cdot \cdots \cdot \|f_n\|_{p(n)}}, \qquad f_j = I(A),\ A \in \mathscr{F}_j;\ f_k \in \mathscr{G}_k\ \forall\, k \neq j,$$

$$\Delta^{(2)} := \sup \frac{|B(f_1, \ldots, f_n)|}{\|f_1\|_{p(1)} \cdot \cdots \cdot \|f_n\|_{p(n)}}, \qquad f_j \in \mathscr{S}(\mathscr{F}_j);\ f_k \in \mathscr{G}_k\ \forall\, k \neq j.$$

Then the following statements hold:

(i) *If $p_j = 1$ then $\Delta^{(2)} = \Delta^{(1)}$.*

(ii) *If $p_j = \infty$ then $\Delta^{(2)} \leqslant 6\Delta^{(1)}$.*

(iii) *If B is a product form and $\sum_{k=1}^n p_k^{-1} \leqslant 1$ then $\Delta^{(2)} \leqslant C_j(\mathbf{p}) \cdot \Delta^{(1)} \cdot (1 - \log \Delta^{(1)})^{1 - 1/p(j)}$, where $C_j(\mathbf{p})$ is a function only of j and $\mathbf{p}$.*

In this lemma, clearly $d_{\mathbf{p}}(B) \leqslant \Delta^{(1)} \leqslant \Delta^{(2)} \leqslant \|B\|_{\mathbf{p}}$. Of course the possibilities $\Delta^{(1)} = 0$ or ∞ may occur, but in these cases the lemma is trivial. Under the hypothesis of (iii) we have $\|B\|_{\mathbf{p}} \leqslant 1$ (as was mentioned earlier) and hence $1 - \log \Delta^{(1)} \geqslant 1$ (in the case $\Delta^{(1)} > 0$). In (iii) of course

$C_j(\mathbf{p})$ need not really depend on j; it can be replaced by $\max_{1 \leqslant j \leqslant n} C_j(\mathbf{p})$. Also, (iii) is redundant if either $p_j = 1$ or $p_j = \infty$.

Before proving Lemma 3.7, let us first quickly show how it can be used to prove Theorems 3.5 and 3.6. In the proofs of both theorems one simply fixes B and $\mathbf{p}$ and then defines Δ_j, $0 \leqslant j \leqslant n$, as follows:

$$\Delta_j := \sup \frac{|B(f_1, \dots, f_n)|}{\|f_1\|_{p(1)} \cdot \cdots \cdot \|f_n\|_{p(n)}}, \qquad f_k \in \mathcal{S}(\mathcal{F}_k) \quad \text{for } k \leqslant j,$$
$$f_k = I(A_k), \qquad A_k \in \mathcal{F}_k, \quad \text{for } k \geqslant j+1.$$

Thus $\Delta_0 = d_{\mathbf{p}}(B)$ and $\Delta_n = \|B\|_{\mathbf{p}}$.

To prove Theorem 3.5(i), set $\mathbf{p} = \infty$ and note that $\Delta_j \leqslant 6\Delta_{j-1}$ $\forall\, j = 1, \dots, n$ by Lemma 3.7(ii). The proofs of Theorem 3.5(ii) and (iii) are similar.

To prove Theorem 3.6 (assuming $d_{\mathbf{p}}(B) > 0$) note that if $p_j = \infty$ then $\Delta_j/\Delta_{j-1} \leqslant 6$, and if $1 \leqslant p_j < \infty$ then

$$\begin{aligned} \Delta_j/\Delta_{j-1} &\leqslant C_j(\mathbf{p}) \cdot (1 - \log \Delta_{j-1})^{1 - 1/p(j)} \\ &\leqslant C_j(\mathbf{p}) \cdot (1 - \log d_{\mathbf{p}}(B))^{1 - 1/p(j)} \end{aligned}$$

(since $d_{\mathbf{p}}(B) \leqslant \Delta_{j-1} \leqslant 1$). Theorem 3.6 follows.

Proof of Lemma 3.7. Without losing generality we assume $j = n$. (Otherwise we can transfer to this case by simply permuting indices.) The proofs of (i) and (ii) are entirely elementary but are given here for completeness.

Proof of (i). Here $p_n = 1$. Let $f_k \in \mathcal{G}_k$ be fixed, $1 \leqslant k \leqslant n-1$. Fix $f_n \in \mathcal{S}(\mathcal{F}_n)$ and represent it by $f_n = \sum_{m=1}^{M} c_m I(A_m)$, where $\{A_1, \dots, A_M\}$ is a partition of Ω_n (with each $A_m \in \mathcal{F}_n$) and $c_m \in \mathbb{C}$ $\forall\, m$. Then

$$\begin{aligned} |B(f_1, \dots, f_n)| &\leqslant \sum_{m=1}^{M} |B(f_1, \dots, f_{n-1}, c_m I(A_m))| \\ &\leqslant \sum_{m=1}^{M} |c_m| \cdot \Delta^{(1)} \cdot \left(\prod_{k=1}^{n-1} \|f_k\|_{p(k)} \right) \cdot P_n(A_m) \\ &= \Delta^{(1)} \cdot \prod_{k=1}^{n} \|f_k\|_{p(k)}. \end{aligned}$$

Part (i) follows.

Proof of (ii). Here $p_n = \infty$. Let $f_1, \dots, f_n$, $\{A_1, \dots, A_M\}$, and $c_1, \dots, c_M$ be as in the proof of (i) above.

Consider first the case where f_n is real, i.e., c_m is real $\forall\, m$. Then

$$\begin{aligned}\operatorname{Re} B(f_1,..., f_n) &= \sum_{m=1}^{M} c_m \operatorname{Re} B(f_1,..., f_{n-1}, I(A_m))\\ &\leqslant \|f_n\|_\infty \cdot \left[\begin{array}{l} \operatorname{Re} B(f_1,..., f_{n-1}, I(F_1)) \\ -\operatorname{Re} B(f_1,..., f_{n-1}, I(F_2)) \end{array}\right]\end{aligned}$$

where

$$F_1 = \bigcup A_m, \qquad \{m: \operatorname{Re} B(f_1,..., f_{n-1}, I(A_m)) \geqslant 0\},$$

$$F_2 = \bigcup A_m, \qquad \{m: \operatorname{Re} B(f_1,..., f_{n-1}, I(A_m)) < 0\}.$$

Hence $\operatorname{Re} B(f_1,..., f_n) \leqslant 2 \cdot \Delta^{(1)} \cdot \prod_{k=1}^{n} \|f_k\|_{p(k)}$. After applying a similar argument to $\operatorname{Re}(-B(f_1,..., f_n))$ we obtain $|\operatorname{Re} B(f_1,..., f_n)| \leqslant 2 \cdot \Delta^{(1)} \cdot \prod_{k=1}^{n} \|f_k\|_{p(k)}$. Similarly $|\operatorname{Im} B(f_1,..., f_n)| \leqslant 2 \cdot \Delta^{(1)} \cdot \prod_{k=1}^{n} \|f_k\|_{p(k)}$. Hence (when f_n is real), $|B(f_1,..., f_n)| \leqslant 8^{1/2} \cdot \Delta^{(1)} \cdot \prod_{k=1}^{n} \|f_k\|_{p(k)}$.

Hence for general $f_n \in \mathscr{S}(\mathscr{F}_n)$,

$$\begin{aligned}|B(f_1,..., f_n)| &\leqslant |B(f_1,..., f_{n-1}, \operatorname{Re} f_n)|\\ &\quad + |B(f_1,..., f_{n-1}, \operatorname{Im} f_n)|\\ &\leqslant 32^{1/2} \cdot \Delta^{(1)} \cdot \prod_{k=1}^{n} \|f_k\|_{p(k)}.\end{aligned}$$

Since $32^{1/2} < 6$, this completes the proof of (ii).

Proof of (iii). For the cases $p_n = 1$ and $p_n = \infty$ one can simply apply (i) and (ii). Now assume $1 < p_n < \infty$. To avoid trivialities, assume $\Delta^{(1)} > 0$ also. Define q, $1 < q < \infty$, by $q^{-1} + p_n^{-1} = 1$. Define $C = C_n(\mathbf{p}) := \max\{6, 6 \cdot C(p_1,..., p_{n-1}; q)\}$, where $C(p_1,..., p_{n-1}; q)$ is taken from Theorem 2.1. (Note that C depends only on $p_1,..., p_n$.) To prove Lemma 3.7(iii) it suffices to prove $\Delta^{(2)} \leqslant C \cdot \Delta^{(1)} \cdot (1 - \log \Delta^{(1)})^{1/q}$.

First, since B is a product form it can be extended to an n-linear form on $\mathscr{S}(\mathscr{F}_1) \times \cdots \times \mathscr{S}(\mathscr{F}_{n-1}) \times \mathscr{L}_1(\mathscr{F}_n)$ retaining the same $\mathbf{i}_n$-norm $\|B\|_{\mathbf{i}(n)} \leqslant 1$. This extended B induces an $(n-1)$-linear operator $T: \mathscr{S}(\mathscr{F}_1) \times \cdots \times \mathscr{S}(\mathscr{F}_{n-1}) \to \mathscr{L}_\infty(\mathscr{F}_n)$ by a well known procedure: For fixed $f_1 \in \mathscr{S}(\mathscr{F}_1),...,$ $f_{n-1} \in \mathscr{S}(\mathscr{F}_{n-1})$ the (extended) mapping $B(f_1,..., f_{n-1}, \cdot)$ is a linear functional on $\mathscr{L}_1(\mathscr{F}_n)$ with norm $\leqslant \prod_{k=1}^{n-1} \|f_k\|_\infty$; hence there is a unique function $g \in \mathscr{L}_\infty(\mathscr{F}_n)$ (depending on $f_1,..., f_{n-1}$) such that $B(f_1,..., f_n) = \int_{\Omega(n)} g f_n \, dP_n$ $\forall\, f_n \in \mathscr{L}_1(\mathscr{F}_n)$, and in fact $\|g\|_\infty \leqslant \prod_{k=1}^{n-1} \|f_k\|_\infty$ since $\|g\|_\infty$ is the norm of the functional; define $T(f_1,..., f_{n-1}) := g$. It is easily seen that T is $(n-1)$-linear, i.e., linear in each coordinate separately. An easy, standard argument shows that T is in fact a product operator.

By Hölder's inequality, $|B(f_1,\ldots,f_n)| \leqslant \|T(f_1,\ldots,f_{n-1})\|_q \cdot \|f_n\|_{p(n)}$ $\forall\,(f_1,\ldots,f_n) \in \mathscr{S}(\mathscr{F}_1) \times \cdots \times \mathscr{S}(\mathscr{F}_n)$. To prove Lemma 3.7(iii) it suffices to show that if $f_k \in \mathscr{G}_k$ $\forall\, k = 1,\ldots, n-1$, then

$$\|T(f_1,\ldots,f_{n-1})\|_q \leqslant C \cdot \Delta^{(1)} \cdot (1 - \log \Delta^{(1)})^{1/q} \cdot \prod_{k=1}^{n-1} \|f_k\|_{p(k)}. \tag{3.1}$$

If $\Delta^{(1)} \geqslant \frac{1}{6}$, then (3.1) holds automatically, since T is a product operator, $C \geqslant 6$, $\Delta^{(1)} \leqslant 1$ (so $\log \Delta^{(1)} \leqslant 0$), and $q^{-1} \geqslant \sum_{k=1}^{n-1} p_k^{-1}$. (See Remark 2.1.)

Henceforth we assume $\Delta^{(1)} < \frac{1}{6}$. We will apply Theorem 2.1, but we first need to establish a version of [37, Lemma 1]:

Claim 0. *If* $t > 0$ *and* $f_k \in \mathscr{G}_k$ $\forall\, k = 1,\ldots, n-1$, *then* $P_n(|T(f_1,\ldots, f_{n-1})| > t) \leqslant [(6\Delta^{(1)}/t) \cdot \prod_{k=1}^{n-1} \|f_k\|_{p(k)}]^q$.

To prove Claim 0, let us first show that

$$P_n(|\mathrm{Re}\, T(f_1,\ldots,f_{n-1})| > 2^{-1/2}t) \leqslant \left[(8^{1/2}\Delta^{(1)}/t) \cdot \prod_{k=1}^{n-1} \|f_k\|_{p(k)}\right]^q. \tag{3.2}$$

We shall simply repeat the argument for [37, Lemma 1] (with trivial modifications). Let $g := T(f_1,\ldots, f_{n-1})$, and define the events (sets) A_1 and $A_2 \subset \Omega_n$ by $A_1 := \{\mathrm{Re}\, g > 2^{-1/2}t\}$ and $A_2 := \{\mathrm{Re}\, g < -2^{-1/2}t\}$. Then for $m = 1, 2$, one has $|\int_{A(m)} g\, dP_n| \geqslant |\int_{A(m)} \mathrm{Re}\, g\, dP_n| \geqslant 2^{-1/2}tP_n(A_m)$, and by our assumptions,

$$\begin{aligned}\left|\int_{A(m)} g\, dP_n\right| &= |B(f_1,\ldots, f_{n-1}, I(A_m))| \\ &\leqslant \Delta^{(1)} \cdot \left(\prod_{k=1}^{n-1} \|f_k\|_{p(k)}\right) \cdot [P_n(A_m)]^{1/p(n)}.\end{aligned}$$

Hence

$$\begin{aligned}[P_n(A_1) + P_n(A_2)] &\leqslant (2^{1/2}/t) \cdot \sum_{m=1}^{2} \left|\int_{A(m)} g\, dP_n\right| \\ &\leqslant (2^{1/2}/t) \cdot \Delta^{(1)} \cdot \left(\prod_{k=1}^{n-1} \|f_k\|_{p(k)}\right) \cdot \sum_{m=1}^{2} [P_n(A_m)]^{1/p(n)}.\end{aligned}$$

Since the last sum is obviously $\leqslant 2[P_n(A_1) + P_n(A_2)]^{1/p(n)}$, we have $[P_n(A_1) + P_n(A_2)]^{1/q} \leqslant (8^{1/2}/t) \cdot \Delta^{(1)} \cdot \prod_{k=1}^{n-1} \|f_k\|_{p(k)}$. Taking both sides to the power q, we get (3.2).

A similar argument gives (3.2) with $\mathrm{Im}\, g$ in place of $\mathrm{Re}\, g$. Combining these two parts, one easily establishes the inequality in Claim 0.

Now we can apply Theorem 2.1, with $\varepsilon := 6\Delta^{(1)} < 1$. Since $1 \leqslant 1 - \log \varepsilon < 1 - \log \Delta^{(1)}$, we have (3.1). This completes the proof of Lemma 3.7.

IV. Measures of Dependence

In this section we shall always be working on a probability space $(\Omega, \mathscr{M}, P)$. Whenever Theorems 3.1–3.6 are applied, it is assumed that the probability space $(\Omega_k, \mathscr{F}_k, P_k)$ in Sections 2–3 satisfy $\Omega_k = \Omega$, $\mathscr{F}_k \subset \mathscr{M}$, and $P_k = P$ (on $\mathscr{F}_k$). Section 4.1 examines measures of dependence between two σ-fields, Section 4.2 examines measures of dependence between three or more σ-fields, Section 4.3 examines measures of dependence involving Hilbert-space-valued r.v.'s, and Section 4.4 gives an exact (sharp) comparison between two particular measures of dependence in a special context.

4.1. *Measures of Dependence between Two σ-Fields*

Let $\mathscr{F}$ and $\mathscr{G}$ be arbitrary but fixed σ-fields ($\subset \mathscr{M}$). Define the well known bilinear form $\mathrm{Cov}\colon \mathscr{S}(\mathscr{F}) \times \mathscr{S}(\mathscr{G}) \to \mathbb{C}$ by $\mathrm{Cov}(f, g) := Efg - EfEg$. (Some people might prefer to replace g by its complex conjugate $\bar{g}$, but that will be of no importance in what follows.) If $(p, q) \in [1, \infty]^2$ then $d_{(p,q)}(\mathrm{Cov}) = \alpha_{1/p,1/q}(\mathscr{F}, \mathscr{G})$. (See (1.1).) For any $(p, q) \in [1, \infty]^2$ define the following measure of dependence:

$$R_{p,q}(\mathscr{F}, \mathscr{G}) := \left[\sup \frac{|\mathrm{Cov}(f, g)|}{\|f\|_p \|g\|_q}, f \in \mathscr{S}(\mathscr{F}), g \in \mathscr{S}(\mathscr{G})\right] = \|\mathrm{Cov}\|_{(p,q)}. \tag{4.1}$$

In this definition we shall not impose the "natural" additional restriction $Ef = Eg = 0$, because the results in Section 3 can be applied more smoothly without it. In any case such a restriction would lower the value by at worst a factor of 1/4, since $\|f - Ef\|_p \leqslant \|f\|_p + |Ef| \leqslant 2\|f\|_p$ holds for every $p \in [1, \infty]$ and every $f \in \mathscr{L}_p(\mathscr{M})$.

Of course in (4.1), if $R_{p,q}(\mathscr{F}, \mathscr{G}) < \infty$ then the same sup is achieved over all $f \in \mathscr{L}_p(\mathscr{F})$ and all $g \in \mathscr{L}_q(\mathscr{G})$.

Also note that $R_{2,2}(\mathscr{F}, \mathscr{G}) = \rho(\mathscr{F}, \mathscr{G})$. (See (1.3).)

The first theorem here is just a list of some results obtained by applying Theorems 3.1–3.6 to the bilinear form Cov. However, for the sake of simplicity we shall not use the full strength of all of these theorems. It should be kept in mind that $d_{(p,q)}(\mathrm{Cov}) \leqslant 1$ whenever $p^{-1} + q^{-1} \leqslant 1$, by the trivial fact that $|P(A \cap B) - P(A)\,P(B)| \leqslant \min\{P(A), P(B)\}$ for any two events A and B. Also $\|\mathrm{Cov}\|_{(p,q)} \leqslant 2$ whenever $p^{-1} + q^{-1} \leqslant 1$. (Theorem 3.6 will be applied to $\frac{1}{2} \cdot \mathrm{Cov}$.) For any $p \in [1, \infty]$, its conjugate exponent will be denoted by p', i.e., $(1/p) + (1/p') = 1$.

THEOREM 4.1. *For any two σ-fields $\mathscr{F}$ and $\mathscr{G}$ the following six statements hold*:

(i) *If* $r \geqslant 0$, $s \geqslant 0$, *and* $r+s<1$, *then* $\alpha_{0,0}(\mathscr{F}, \mathscr{G}) \leqslant \alpha_{r,s}(\mathscr{F}, \mathscr{G}) \leqslant [\alpha_{0,0}(\mathscr{F}, \mathscr{G})]^{1-r-s}$.

(ii) *If* $0<r<s<1$ *then* $\alpha_{r,1-r}(\mathscr{F}, \mathscr{G}) \leqslant [\alpha_{s,1-s}(\mathscr{F}, \mathscr{G})]^{r/s}$ *and* $\alpha_{s,1-s}(\mathscr{F}, \mathscr{G}) \leqslant [\alpha_{r,1-r}(\mathscr{F}, \mathscr{G})]^{(1-s)/(1-r)}$.

(iii) *If* $1<p$, $q \leqslant \infty$ *with* $p^{-1}+q^{-1}<1$, *then* $R_{\infty,\infty}(\mathscr{F}, \mathscr{G}) \leqslant R_{p,q}(\mathscr{F}, \mathscr{G}) \leqslant 2^{(1/p)+(1/q)}[R_{\infty,\infty}(\mathscr{F}, \mathscr{G})]^{1-(1/p)-(1/q)}$.

(iv) *If* $1<p<q<\infty$ *then* $R_{p,p'}(\mathscr{F}, \mathscr{G}) \leqslant 2^{1-q'/p'} \cdot [R_{q,q'}(\mathscr{F}, \mathscr{G})]^{q'/p'}$ *and* $R_{q,q'}(\mathscr{F}, \mathscr{G}) \leqslant 2^{1-p/q} \cdot [R_{p,p'}(\mathscr{F}, \mathscr{G})]^{p/q}$.

(v) $R_{\infty,\infty}(\mathscr{F}, \mathscr{G}) \leqslant 36\alpha_{0,0}(\mathscr{F}, \mathscr{G})$, $R_{1,\infty}(\mathscr{F}, \mathscr{G}) \leqslant 6\alpha_{1,0}(\mathscr{F}, \mathscr{G})$, $R_{\infty,1}(\mathscr{F}, \mathscr{G}) \leqslant 6\alpha_{0,1}(\mathscr{F}, \mathscr{G})$, *and* $R_{1,1}(\mathscr{F}, \mathscr{G}) = \alpha_{1,1}(\mathscr{F}, \mathscr{G})$.

(vi) *If* $1 \leqslant p, q \leqslant \infty$ *and* $p^{-1}+q^{-1} \leqslant 1$, *then* $R_{p,q}(\mathscr{F}, \mathscr{G}) \leqslant C \cdot \alpha_{1/p,1/q}(\mathscr{F}, \mathscr{G}) \cdot [1-\log \alpha_{1/p,1/q}(\mathscr{F}, \mathscr{G})]^c$, *where the constants* $C=C(p,q)$ *and* $c=(p,q)$ *are functions only of p and q, the latter constant being as follows*: $c(\infty,\infty)=0$, $c(p,\infty)=1-p^{-1}$ *if* $1 \leqslant p<\infty$, $c(\infty,q)=1-q^{-1}$ *if* $1 \leqslant q<\infty$, *and* $c(p,q)=2-(p^{-1}+q^{-1})$ *if* $1<p$, $q<\infty$.

Statement (iv) was given (in the language of linear operators) by Rosenblatt [33, Chap. 7]. Dvoretzky [12, p. 528, Lemma 5.3] showed that in the first inequality in (v) the 36 can be replaced by the much better constant 2π. The last equality in (v) appeared in [2]. The other two inequalities in (v) are also well known. Statement (vi) sharpens the main result in [3].

Remark 4.1. In more or less the same spirit as in Remark 3.2, one might regard two measures of dependence as "equivalent" if each one becomes arbitrarily small as the other becomes sufficiently small. In this sense, by Theorem 4.1, (i) the dependence coefficients $\alpha_{r,s}$, $0 \leqslant r$, $s<1$, $r+s<1$, and $R_{p,q}$, $1<p$, $q \leqslant \infty$, $p^{-1}+q^{-1}<1$, are all equivalent to each other; and (ii) the dependence coefficients $\alpha_{r,1-r}$, $0<r<1$, and $R_{p,p'}$, $1<p<\infty$, are all equivalent to each other. This understanding may be helpful in trying to fit into a comprehensible structure the numerous measures of dependence that have been studied in the literature. It appears that many of them belong to one of the four distinct equivalence classes represented by $\alpha_{0,0}$, $\alpha_{1,0}$, $\alpha_{0,1}$, and $\alpha_{1/2,1/2}$; and so these four equivalence classes would perhaps be a "central" part of such a structure. These four "central" classes—as we shall call them here for convenience—correspond to four equivalence classes of mixing conditions on Markov chains that were discussed by Rosenblatt [33, Chap. 7]. The measures of dependence based on Hilbert-space-valued r.v.'s that will be discussed later on in Section 4.3, also belong to these four "central" classes (by Theorem 4.2 in that section). Also, the dependence coefficient $b_p(\cdot, \cdot)$ defined in [4] by

$$b_p(\mathscr{F}, \mathscr{G}) := \sup \|P(A \mid \mathscr{G}) - P(A)\|_p, \qquad A \in \mathscr{F}$$

is equivalent to $\alpha_{0,0}$ (resp. $\alpha_{0,1}$) if $1 \leqslant p < \infty$ (resp. if $p = \infty$). Of course, two very important measures of dependence that do not belong to the four "central" classes are $\alpha_{1,1}$ and the one that is the basis for the "absolute regularity" condition (see Volkonskii and Rozanov [38]). Whatever "equivalence structure" there is for the measures of dependence $\alpha_{r,s}$, $r+s>1$, and $R_{p,q}$, $p^{-1}+q^{-1}>1$, seems to be somewhat complicated and not so easy to decipher; it will not be treated further here.

Proof of Theorem 1.1. To prove (i), simply note that by Theorems 3.3 and 3.5 and [12, Lemma 5.3],

$$\begin{aligned} R_{p,q}(\mathscr{F}, \mathscr{G}) &\leqslant [R_{\infty,\infty}(\mathscr{F}, \mathscr{G})]^{1/t} \cdot [R_{1,\infty}(\mathscr{F}, \mathscr{G})]^{1/p} \cdot [R_{\infty,1}(\mathscr{F}, \mathscr{G})]^{1/q} \\ &\leqslant [2\pi\alpha_{0,0}(\mathscr{F}, \mathscr{G})]^{1/t} \cdot [6\alpha_{1,0}(\mathscr{F}, \mathscr{G})]^{1/p} \cdot [6\alpha_{0,1}(\mathscr{F}, \mathscr{G})]^{1/q}. \end{aligned}$$

To prove (ii), first assume without loss of generality that $p \leqslant q$ (and hence $p \leqslant 2 \leqslant q$ by our assumption $p^{-1}+q^{-1}=1$). The case where $p=1$ (and $q=\infty$) is trivial, so let us assume $p>1$. By Theorem 4.1(iv),(vi),

$$\begin{aligned} R_{p,q}(\mathscr{F}, \mathscr{G}) &\leqslant 2^{1-2/q}[R_{2,2}(\mathscr{F}, \mathscr{G})]^{2/q} \\ &\leqslant 2^{1-2/q} \cdot [C(2,2) \cdot \alpha_{1/2,1/2}(\mathscr{F}, \mathscr{G}) \cdot [1 - \log \alpha_{1/2,1/2}(\mathscr{F}, \mathscr{G})]]^{2/q}. \end{aligned}$$

The upper bound of 3000 on $C(2, 2)$ (a rather crude bound) can be seen as follows: By Theorem 2.2 one can take $C(2; 2) = 6$ in Theorem 2.1. From the proof of Lemma 3.7(iii) we see that in that lemma one can take $C_1((2, 2)) = C_2((2, 2)) = 6^2$. From the proof of Theorem 3.6 we see that one can take $C((2, 2)) = 6^4$ there. Since Theorem 4.1(vi) was obtained by applying Theorem 3.6 to $\frac{1}{2} \cdot \mathrm{Cov}$, a simple, crude calculation shows that in Theorem 4.1(vi) one can take $C(2, 2) = 2 \cdot 6^4 < 3000$. This completes the proof of Theorem 1.1.

Remark 4.2. The upper bound of 3000 on $C(2, 2)$ can be substantially lowered by the following approach: Use [39, p. 512, Theorem 1.1] to convert this task into one involving just *real*-valued functions; take advantage of this in order to lower the constant 6 in Claim 0 in the proof of Lemma 3.7(iii) (look at the original argument in [37, Lemma 1]); and at an appropriate place in the proof of Theorem 3.6, use the sharp result (Theorem 4.3) in Section 4.4. The details are left to the reader.

4.2. *Measures of Dependence between Three or More σ-Fields*

Measures of dependence between three or more σ-fields have been studied by several authors, e.g., Statulevičius [35, 36], Žurbenko [41], Mase [25], and Dmitrovskii, Ermakov, and Ostrovskii [11]. Some of their results can, except perhaps for a constant factor, be derived from Theorems 3.1–3.6. This was part of the motivation for presenting Theorems 3.1–3.6 in their present (multidimensional) form instead of limiting them to the case $n=2$.

Let us consider an example of Mase [25]. Suppose $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$ are sub-σ-fields of $\mathcal{M}$. Mase considered the 4-linear form "Cum" (cumulant) on, say, $\mathcal{S}(\mathcal{M}_1)\times\mathcal{S}(\mathcal{M}_2)\times\mathcal{S}(\mathcal{M}_3)\times\mathcal{S}(\mathcal{M}_4)$, defined by

$$\begin{aligned}\mathrm{Cum}(f_1, f_2, f_3, f_4) &:= E\prod_{k=1}^{4}(f_k - Ef_k)\\ &\quad -\sum\,[E(f_a - Ef_a)(f_b - Ef_b)]\cdot[E(f_c - Ef_c)(f_d - Ef_d)]\end{aligned}$$

where the sum is taken over $(a, b, c, d)\in\{(1, 2, 3, 4), (1, 3, 2, 4), (1, 4, 2, 3)\}$. (We retain this definition of "cumulant" even when the f_k's are complex-valued.) Then for a given $\delta>0$, Theorem 1 of [25] can be obtained from the following inequalities, where $\mathbf{p}=(4+\delta, 4+\delta, 4+\delta, 4+\delta)$:

$$\begin{aligned}\|\mathrm{Cum}\|_{\mathbf{p}} &\leqslant [\|\mathrm{Cum}\|_{\infty}]^{\delta/(4+\delta)}\cdot\prod_{k=1}^{4}[\|\mathrm{Cum}\|_{\mathbf{i}(k)}]^{1/(4+\delta)}\\ &\leqslant [6^{4\delta/(4+\delta)}\cdot 64^{4/(4+\delta)}]\cdot[d_{\infty}(\mathrm{Cum})]^{\delta/(4+\delta)}.\end{aligned}$$

Here the first inequality comes from Theorem 3.3, and the second from Theorem 3.5(i) and the elementary fact that $\|\mathrm{Cum}\|_{\mathbf{i}(k)}\leqslant 64$ $\forall\, k=1, 2, 3, 4$.

Here is another possible example. By Theorem 3.6 (applied to $(1/64)\cdot\mathrm{Cum}$), there exists a constant K such that the inequality $\|\mathrm{Cum}\|_{(4,4,4,4)}\leqslant K\cdot d_{(4,4,4,4)}(\mathrm{Cum})\cdot[1-\log d_{(4,4,4,4)}(\mathrm{Cum})]^3$ always holds. Such an inequality might be useful, for example, in verifying an assumption such as $\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}|\mathrm{Cum}(X_0, X_i, X_j, X_k)|<\infty$ for strictly stationary sequences of *real*-valued r.v.'s $(X_k,\ k=..., -1, 0, 1,...)$ under certain dependence assumptions. This latter inequality has played a natural role in certain kinds of limit theorems, especially for estimators of parameters in time series analysis; see e.g., Hannan [15, pp. 226–227] and Mase [25, Theorem 3].

Some general comments might be worth making. Suppose $n\geqslant 2$ and $\mathcal{M}_1,..., \mathcal{M}_n$ are σ-fields ($\subset\mathcal{M}$). If $\mathcal{P}=\{S_1, S_2,..., S_M\}$ is a partition of $\{1, 2,..., n\}$ with each S_m non-empty, then the n-linear form

$B_{\mathscr{P}}(f_1,..., f_n) := \prod_{m=1}^{M} E(\prod_{k \in S(m)} f_k)$, defined on $\mathscr{S}(\mathscr{M}_1) \times \cdots \times \mathscr{S}(\mathscr{M}_n)$, is a product form by the multidimensional Hölder inequality. If $B_0 = \sum_{j=1}^{J} c_j B_{\mathscr{P}(j)}$, where $\mathscr{P}(1),..., \mathscr{P}(J)$ are each a partition of $\{1,..., n\}$ and $c_1,..., c_J$ are complex numbers, then $c^{-1}B_0$ is a product form for $c = \sum_{j=1}^{J} |c_j|$. For appropriate choices of such n-linear forms B_0, one might regard the norms $d_{\mathbf{p}}(B_0)$ and $\|B_0\|_{\mathbf{p}}$ for vectors $\mathbf{p} \in [1, \infty]^n$, as measures of "multidimensional" dependence between the σ-fields $\mathscr{G}_1,..., \mathscr{G}_n$. By applying Theorems 3.1–3.6 to appropriate forms B_0 (or to $c^{-1}B_0$ when a product form is called for), one can obtain numerous inequalities, including ones analogous to Theorem 1.1(i),(ii) and Theorem 4.1. In order to regard $d_{\mathbf{p}}(B_0)$ or $\|B_0\|_{\mathbf{p}}$ as a "measure of dependence," one would naturally want B_0 to be such that $d_\infty(B_0) = 0$ whenever $\mathscr{G}_1,..., \mathscr{G}_n$ are independent σ-fields. (The forms "Cov" and "Cum" used above both have this property.) Of course even for a "natural" form B_0, only certain kinds of dependence might be detected. For example, if $n = 4$, $\mathscr{G}_1$, $\mathscr{G}_2$, and $\mathscr{G}_3$ are independent σ-fields, and $\mathscr{G}_4 = \mathscr{G}_3$, then $d_\infty(\text{Cum}) = 0$ and thus the (severe) dependence between $\mathscr{G}_3$ and $\mathscr{G}_4$ is not detected by the dependence coefficient $d_\infty(\text{Cum})$.

Without getting into the details, here are a few more possibilities for applications of interpolation theory to measures of dependence:

(i) Theorem 3.1 (resp. Theorem 3.2) is in an obvious way an analog of Theorem 3.3 (resp. Theorem 3.4). In our present context there are other applications of the (multidimensional) Riesz–Thorin interpolation theorem that are similar to but outside of the narrow scope of Theorems 3.3–3.4, and they have analogs similar to but outside of Theorems 3.1–3.2. A wider class of inequalities for measures of dependence can be derived if one makes use of these more general applications (in addition to Theorems 3.1–3.6).

(ii) Interpolation theory on Orlicz spaces (see, e.g., Peetre [27], Gustavsson and Peetre [14], and the references therein) might be useful in obtaining inequalities such as the ones by Zuev [40] and Žurbenko and Zuev [42] alluded to in Remark 1.1. The (essentially unavoidable) log terms in those inequalities might well be due to the same basic underlying "cause" as the log terms in Theorems 1.1, 2.1, 3.6, and 4.1.

(iii) An operator $T(f)$ is called "sublinear" if there is a constant K such that the inequality $|T(f+g)| \leqslant K(|T(f)| + |T(g)|)$ always holds. Similarly an operator $T(f_1,..., f_n)$ is called "multi-sub-linear" if it is sublinear in each coordinate separately. It appears that Theorem 2.1 can perhaps be extended to some multi-sub-linear operators (besides multilinear ones), and Theorem 3.6 to some multi-sub-linear forms. This would perhaps lead to inequalities, such as ones similar to Theorems 1.1 and 4.1, for a broader class of measures of dependence (between two or more σ-fields) than the measures discussed here.

4.3. *Measures of Dependence Involving H-Valued r.v.'s*

Because of the research that has been done on limit theorems for dependent sequences of Hilbert-space-valued (H-valued) random variables (see, e.g., [8, 24]), it might be worthwhile to take a quick look at measures of dependence involving H-valued r.v.'s and see how they relate to the other measures of dependence discussed so far. We shall follow the basic approach of Dehling and Philipp [8], where a well known theorem of Grothendieck (see [23, p. 68]) is used in order to derive "moment inequalities" for H-valued random variables. Just for simplicity, we shall restrict our attention to real (not complex) Hilbert spaces H and to mean-zero H-valued (strongly measurable) r.v.'s.

Let H be an arbitrary real Hilbert space, and let the inner product be denoted by $(\cdot, \cdot)$. For any two σ-fields $\mathscr{F}$ and $\mathscr{G} \subset \mathscr{M}$ and any vector $(p, q) \in [1, \infty]^2$ with $p^{-1} + q^{-1} \leqslant 1$, define the measure of dependence

$$\rho^H_{p,q}(\mathscr{F}, \mathscr{G}) := \sup \frac{|E(f, g)|}{\|f\|_p \, \|g\|_q}$$

where this sup is taken over all H-valued r.v.'s f and g such that f is $\mathscr{F}$-measurable, g is $\mathscr{G}$-measurable, $\|f\|_p < \infty$, $\|g\|_q < \infty$, and $Ef = Eg = 0$. Here of course $\|X\|_p := \|(X, X)^{1/2}\|_p$ for any H-valued r.v. X and any $p \in [1, \infty]$. Note that $R^H_{p,q}(\mathscr{F}, \mathscr{G}) \leqslant 16 \cdot \rho^H_{p,q}(\mathscr{F}, \mathscr{G})$. (The 16 could of course be omitted if we revised (4.1) by taking only real, mean-zero r.v.'s into account.)

THEOREM 4.2. *If H is a real Hilbert space, $(p, q) \in [1, \infty]^2$ with $p^{-1} + q^{-1} \leqslant 1$, and $\mathscr{F}, \mathscr{G} \subset \mathscr{M}$, then $\rho^H_{p,q}(\mathscr{F}, \mathscr{G}) \leqslant A \cdot R_{p,q}(\mathscr{F}, \mathscr{G})$, where $A = A(p, q)$ is a function only of p and q. Moreover $\rho^H_{2,2}(\mathscr{F}, \mathscr{G}) = \rho(\mathscr{F}, \mathscr{G}) = R_{2,2}(\mathscr{F}, \mathscr{G})$.*

Theorem 4.2 is not new; it is simply a formulation, in our context, of results that are well known in other contexts (e.g., in functional analysis). Its presentation here is motivated partly by Dehling and Philipp [8, Lemma 2.2], in which this theorem for the case $p = q = \infty$ was shown to follow from Grothendieck's inequality, and partly by a simple proof of the very last statement ($\rho^H_{2,2}(\mathscr{F}, \mathscr{G}) = \rho(\mathscr{F}, \mathscr{G})$) that was shown to one of the authors by S. Kwapień. The cases $(p, q) \neq (\infty, \infty)$ can be made very transparent with totally elementary arguments. Although these arguments are well known in various forms (see, e.g., Pietsch [29, Chap. 22]), they will be repeated here in our terminology for the sake of expository clarity.

For $1 \leqslant p < \infty$ define $Z_p := [(2\pi)^{-1/2} \int_{-\infty}^{\infty} |x|^p e^{-x^2/2} \, dx]^{1/p}$, the p-norm of a real $N(0, 1)$ r.v. We shall consider just these two cases:

Case I. $1 \leqslant p < \infty$ and $q = \infty$, with $A = A(p, \infty) = (\pi/2)^{1/2} Z_p$.

Case II. $1 < p, q < \infty$ with $p^{-1} + q^{-1} \leqslant 1$, with $A = A(p, q) = Z_p \cdot Z_q$.
(Then the very last part of Theorem 4.2 becomes a simple consequence of (1.3), Case II, and the fact $Z_2 = 1$.) We shall use a Gaussian measure on H as in Rietz [30].

Proof for Case I. By a standard approximation argument it suffices to consider a finite dimensional Hilbert space H. Let γ be a standard Gaussian measure on H, i.e., γ is the distribution of the r.v. $\sum_{i=1}^{\dim(H)} \gamma_i e_i$, where $e_1, e_2, \ldots, e_{\dim(H)}$ is an orthonormal basis for H and the γ_i's are independent real $N(0, 1)$ r.v.'s.

LEMMA O. $\forall\, x, y \in H$,

$$(x, y) = |y| \cdot (\pi/2)^{1/2} \cdot \int_H (x, u) \operatorname{sign}(y, u)\, \gamma(du).$$

Here $|y|$ *just means* $(y, y)^{1/2}$.

Thanks to the spherical symmetry of the measure γ, to prove Lemma O it suffices to consider the case $y = e_1$. For $x = e_1$, Lemma O simply boils down to the fact $E|\gamma_1| = (2/\pi)^{1/2}$, an elementary property of $N(0, 1)$ r.v.'s. For $x = e_j$, $j \neq 1$, Lemma O boils down to $E\gamma_j \operatorname{sign}(\gamma_1) = 0$. And now Lemma O can be established for general x by the usual equation $x = \sum_i (x, e_i)\, e_i$.

Now to prove Case I let f and g be arbitrary H-valued r.v.'s, measurable with respect to $\mathscr{F}$ and $\mathscr{G}$, respectively, such that $\|f\|_p < \infty$, $\|g\|_\infty < \infty$, and $Ef = Eg = 0$. Using Lemma O and Fubini's theorem, we have

$$\begin{aligned} E(f, g) &= (\pi/2)^{1/2} \int_H E[(f, u) \cdot \operatorname{sign}(g, u) \cdot |g|]\, \gamma(du) \\ &\leqslant (\pi/2)^{1/2} \int_H R_{(p,\infty)}(\mathscr{F}, \mathscr{G}) \cdot \|g\|_\infty \cdot (E|(f, u)|^p)^{1/p}\, \gamma(du) \\ &\leqslant (\pi/2)^{1/2} R_{(p,\infty)}(\mathscr{F}, \mathscr{G}) \cdot \|g\|_\infty \cdot \left[E \int_H |(f, u)|^p\, \gamma(du) \right]^{1/p} \\ &= (\pi/2)^{1/2} R_{(p,\infty)}(\mathscr{F}, \mathscr{G}) \cdot \|g\|_\infty \cdot \|f\|_p \cdot Z_p \end{aligned}$$

where, in order to obtain the last equality, we use the following elementary fact: If $y \in H$ and $|y| = 1$ then on the probability space (H, γ) the r.v. $Y\colon H \to \mathbb{R}$ defined by $Y(u) = (y, u)$ has the $N(0, 1)$ distribution. Case I is proved.

Proof for Case II. Use the elementary identity $(x, y) = \int_H (x, u) \cdot (y, u)\, \gamma(du)$ in place of Lemma O. (It can be proved in the same way as Lemma O.) The rest of the argument is essentially the same as for Case I.

Remark 4.3. Theorem 1.1 now extends to H-valued r.v.'s X and Y, with $|E(X, Y) - (EX, EY)|$ in place of $|EXY - EXEY|$. If H is a real Hilbert space and $EX = EY = 0$, then it suffices to multiply the factors 2π and 3000 in Theorem 1.1 by $A(p, q)$ from Theorem 4.2. If one wishes to allow H to be a complex Hilbert space and remove the restriction $EX = EY = 0$, then an additional constant factor may be needed.

4.4. *An Exact Comparison*

Referring to Theorems 3.6 and 4.1(vi), we shall present a special situation in which a closely related exact (sharp) inequality has been established. First, for any two σ-fields $\mathscr{F}$ and $\mathscr{G}$, define the measure of dependence

$$\tau(\mathscr{F}, \mathscr{G}) := \sup |\mathrm{Corr}(I(A), I(B))|, \qquad A \in \mathscr{F},\ B \in \mathscr{G}.$$

Now $|\mathrm{Corr}(I(A), I(B))|$ and $|P(A \cap B) - P(A)\, P(B)|$ each remain unchanged if A (or B) is replaced by its complement. Hence, in both the definition of $\tau(\cdot, \cdot)$ and of $\lambda(\cdot, \cdot)$ (see (1.2)) the sup can be taken over events A and B with probability $\leqslant \frac{1}{2}$. It follows immediately that

$$\lambda(\mathscr{F}, \mathscr{G}) \leqslant \tau(\mathscr{F}, \mathscr{G}) \leqslant 2\lambda(\mathscr{F}, \mathscr{G}). \tag{4.2}$$

(This inequality was noted in [3].)

THEOREM 4.3. *Suppose $\mathscr{F} = \{\Omega, \phi, A, A^c\}$ for some event A, and $\mathscr{G}$ is any σ-field. Then $\rho(\mathscr{F}, \mathscr{G}) \leqslant \tau(\mathscr{F}, \mathscr{G}) \cdot [1 - \log \tau(\mathscr{F}, \mathscr{G})]^{1/2}$.*

Here of course A^c denotes the complement of A, and $\rho(\mathscr{F}, \mathscr{G})$ is the maximal correlation (see (1.3)). In Example 4.4 below it will be shown that this inequality is sharp.

Proof of Theorem 4.3. Let $t = \tau(\mathscr{F}, \mathscr{G})$. If $t = 0$ or 1 then Theorem 4.3 is trivial. So we assume $0 < t < 1$. (Of course $t > 0$ implies $0 < P(A) < 1$.)

Let f and g be of the form

$$f = f_1 I(A) + f_2 I(A^c), \qquad g = \sum_{j=1}^{J} g_j I(B_j)$$

where $J \geqslant 2$, $\{B_1, ..., B_J\}$ is a $\mathscr{G}$-measurable partition of Ω, $P(B_j) > 0\ \forall j$, $f_1 < f_2$, $g_1 < g_2 < \cdots < g_J$, and $Ef = Eg = 0$. It suffices to prove

$|\mathrm{Corr}(f, g)| \leqslant t \cdot (1 - \log t)^{1/2}$. Since $-g$ can be expressed in the same way as g, it in fact suffices to prove

$$\mathrm{Corr}(f, g) \leqslant t \cdot (1 - \log t)^{1/2} \tag{4.3}$$

Similarly to the argument in [3], define the positive numbers $q = f_2 - f_1$, $c = P(A)$, and the r.v. $V = c - I(A)$; and for each $j = 1, ..., J-1$ define the event $D_j = \bigcup_{k=1}^{j} B_k$, the positive numbers $r_j = g_{j+1} - g_j$ and $d_j = P(D_j)$, and the r.v. $W_j = d_j - I(D_j)$. Note that $EV = 0$ and $EW_j = 0\ \forall\, j$. Keeping in mind that $Ef = Eg = 0$ by assumption, one can easily show that

$$f = qV, \qquad g = \sum_{j=1}^{J-1} r_j W_j. \tag{4.4}$$

We need to digress for a moment to define several functions on the unit interval. For each x, $0 < x < 1$, define the function $H_x(\cdot)$ on $[0, 1]$ as

$$H_x(y) := \min\{x(1-y),\, y(1-x),\, t[x(1-x)\, y(1-y)]^{1/2}\} \tag{4.5}$$

and define the numbers

$$\gamma_x := \frac{t^2 x}{(1-x) + t^2 x}, \qquad \mu_x := \frac{x}{x + t^2(1-x)}. \tag{4.6}$$

With a little arithmetic one can show that for $0 < x < 1$ one has

$$\begin{aligned} &0 < \gamma_x < x < \mu_x < 1, \\ H_x(y) &= y(1-x) && \text{if } 0 \leqslant y \leqslant \gamma_x \\ &= t[x(1-x)\, y(1-y)]^{1/2} && \text{if } \gamma_x \leqslant y \leqslant \mu_x \\ &= x(1-y) && \text{if } \mu_x \leqslant y \leqslant 1. \end{aligned} \tag{4.7}$$

For each $0 < x < 1$ define the function $h_x(\cdot)$ on $[0, 1]$ by

$$\begin{aligned} h_x(y) &= 1 - x && \text{if } 0 \leqslant y \leqslant \gamma_x \\ &= \frac{t[x(1-x)]^{1/2}(1-2y)}{2[y(1-y)]^{1/2}} && \text{if } \gamma_x < y < \mu_x \\ &= -x && \text{if } \mu_x \leqslant y \leqslant 1. \end{aligned} \tag{4.8}$$

For each fixed x, $0 < x < 1$, $H_x(y)$ is continuous, $h_x(y)$ is bounded and monotonically decreasing and (except at $y = 0$, γ_x, μ_x, and 1) is the derivative of H_x, and so H_x is an indefinite integral of h_x on $[0, 1]$.

Define the numbers $d_0 = 0$ and $d_J = 1$ and define the function g^* on the half-open unit interval $(0, 1]$ by

$$g^*(y) := \sum_{j=1}^{J} g_j I_{(d_{j-1}, d_j]}(y).$$

Now for each j, $\mathrm{Corr}(V, W_j) = \mathrm{Corr}(I(A), I(D_j)) \leqslant t$, and by taking into account the additional restriction $P(A \cap D_j) \leqslant \min\{P(A), P(D_j)\}$ we find that $EVW_j \leqslant H_c(d_j)$ by (4.5). Keeping in mind that $H_x(0) = H_x(1) = 0$ for $0 < x < 1$ and q, r_j are positive, we have by (4.4)

$$\begin{aligned} Efg &\leqslant \sum_{j=1}^{J-1} q r_j H_c(d_j) = q \sum_{j=1}^{J} g_j [H_c(d_{j-1}) - H_c(d_j)] \\ &= q \sum_{j=1}^{J} g_j \int_{d(j-1)}^{d(j)} h_c(y)\, dy \\ &= \int_0^1 q \cdot g^*(y) \cdot h_c(y)\, dy. \end{aligned}$$

Now $Eg^2 = \int_0^1 (g^*(y))^2\, dy$. Thus by the Cauchy–Schwarz inequality, $Efg \leqslant (Eg^2)^{1/2} (\int_0^1 q^2 (h_c(y))^2\, dy)^{1/2}$. Also, $Ef^2 = q^2 c(1-c)$. Hence, to verify (4.3) and thereby prove Theorem 4.3 it suffices to prove that

$$\int_0^1 h_c^2(y)\, dy = c(1-c)\, t^2 (1 - \log t). \tag{4.9}$$

Using (4.8), the integral is equal to

$$(1-c)^2 \gamma_c + t^2 c(1-c) \int_{\gamma_c}^{\mu_c} \frac{(1-2y)^2}{4y(1-y)}\, dy + c^2 (1 - \mu_c).$$

Now $(\log y - \log(1-y) - 4y)$ is an indefinite integral of $(1-2y)^2/(y(1-y))$, and hence by (4.6) and some arithmetic, (4.9) follows. This completes the proof of Theorem 4.3.

EXAMPLE 4.4. With this example we shall show that Theorem 4.3 is sharp no matter what $P(A)$ is, as long as $0 < P(A) < 1$. Let $0 < s < 1$ and $0 < t < 1$ be fixed. We shall construct a probability space $(\Omega, \mathscr{M}, P)$ and σ-fields $\mathscr{F} = (\Omega, \phi, A, A^c)$ and $\mathscr{G} \subset \mathscr{M}$ such that $P(A) = s$, $\tau(\mathscr{F}, \mathscr{G}) = t$, and $\rho(\mathscr{F}, \mathscr{G}) = t \cdot (1 - \log t)^{1/2}$.

Let $\Omega = \{0, 1\} \times [0, 1]$ (the union of two line segments in the plane), let $\mathscr{M}$ be the family of all Borel subsets of Ω, and define the random variables X and Y by $X(x, y) := x$ and $Y(x, y) := y$ for $(x, y) \in \Omega$. Let P be the probability measure on $(\Omega, \mathscr{M})$ determined by the following condition: for each Borel subset $B \subset [0, 1]$,

$$P(X=0,\ Y \in B) = \int_B (h_s(y) + s)\, dy,$$

$$P(X=1,\ Y \in B) = \int_B (1 - s - h_s(y))\, dy$$

where $h_s(y)$ is as defined in (4.8) (in the proof of Theorem 4.3); it was noted in the proof of Theorem 4.3 that $h_s(y)$ is monotonically decreasing on $[0, 1]$, so that $-s \leqslant h_s(y) \leqslant 1 - s\ \forall\ y \in [0, 1]$, and it follows that both integrands here are always non-negative. Define the σ-fields $\mathscr{F} = \sigma(X)$ and $\mathscr{G} = \sigma(Y)$. Defining the event $A = \{X = 0\}$ we have $\mathscr{F} = \{\Omega, \phi, A, A^c\}$ and $P(A) = s$. Also, Y is uniformly distributed on $[0, 1]$.

We shall first show that $\tau(\mathscr{F}, \mathscr{G}) \leqslant t$. Let B be any Borel subset of the interval $[0, 1]$ and let $b = P(Y \in B) =$ the Lebesgue measure of B. Since $\operatorname{Corr}(I(A), I(Y \in B))$ simply changes sign if A is replaced by A^c or B by B^c, it suffices to prove $\operatorname{Corr}(I(A), I(Y \in B)) \leqslant t$. Since $h_s(y)$ is monotonically decreasing,

$$\begin{aligned} P(X=0,\ Y \in B) &= sb + \int_B h_s(y)\, dy \leqslant sb + \int_0^b h_s(y)\, dy \\ &= sb + H_s(b) \leqslant sb + t[s(1-s)\, b(1-b)]^{1/2} \qquad (4.10) \end{aligned}$$

by (4.5). (Recall that H_s is an indefinite integral of h_s.) Since $P(X=0) = s$ and $P(Y \in B) = b$, (4.10) implies $\operatorname{Corr}(I(A), I(Y \in B)) \leqslant t$. Hence $\tau(\mathscr{F}, \mathscr{G}) \leqslant t$.

To show that $\tau(\mathscr{F}, \mathscr{G}) = t$, note that both inequalities in (4.10) become equalities if we let $B = [0, b]$ with $\gamma_s \leqslant b \leqslant \mu_s$ (see (4.7)), and in this case $\operatorname{Corr}(I(A), I(Y \in B)) = t$.

Finally we need to show that $\rho(\mathscr{F}, \mathscr{G}) = t \cdot (1 - \log t)^{1/2}$. Define the random variable Z by $Z = h_s(Y)$; then Z is $\mathscr{G}$-measurable. Note that $EZ = \int_0^1 h_s(y)\, dy = 0$, $EZ^2 = \int_0^1 h_s^2(y)\, dy$, and $EI(A)\, Z = \int_0^1 h_s(y)(h_s(y) + s)\, dy = EZ^2$ by the definition of P. Hence

$$\begin{aligned} \operatorname{Corr}(I(A), Z) &= [(EZ^2)/(\operatorname{Var} I(A))]^{1/2} \\ &= [(EZ^2)/(s(1-s))]^{1/2} = [t^2(1 - \log t)]^{1/2} \end{aligned}$$

by (4.9). Hence $\rho(\mathscr{F}, \mathscr{G}) = t(1 - \log t)^{1/2}$ (by Theorem 4.3). This completes Example 4.4.

Remark 4.4. Example 4.4 shows that for the vector $(\infty, 2)$, Theorem 4.1(vi) (as well as Theorem 3.6) is within a constant factor of being sharp. Simply set $s = \frac{1}{2}$ in Example 4.4 (so that $P(A) = P(A^c) = 1/2$) and observe that $\alpha_{0,1/2}(\mathscr{F}, \mathscr{G}) \leqslant \lambda(\mathscr{F}, \mathscr{G}) \leqslant \tau(\mathscr{F}, \mathscr{G})$ and $R_{\infty,2}(\mathscr{F}, \mathscr{G}) \geqslant 2^{-1/2} R_{2,2}(\mathscr{F}, \mathscr{G}) = 2^{-1/2} \rho(\mathscr{F}, \mathscr{G})$.

Remark 4.5. In the special case when $n = 2$ and $p_1 = q = 2$, Theorem 2.1 is within a constant factor of being sharp. Let ε, $0 < \varepsilon < 1$, be arbitrary but fixed. Using Example 4.4, let $(\Omega, \mathscr{M}, P)$ be a probability space with σ-fields $\mathscr{M}_1$ and $\mathscr{M}_2 \subset \mathscr{M}$ such that $\tau(\mathscr{M}_1, \mathscr{M}_2) = \varepsilon/3$, $\rho(\mathscr{M}_1, \mathscr{M}_2) = (\varepsilon/3) \cdot [1 - \log(\varepsilon/3)]^{1/2}$, and $\mathscr{M}_1 = \{\Omega, \phi, A, A^c\}$ for some event A. Next let us turn our attention to Lemma 3.7(iii) in the case where $n = 2$, $j = 2$, $p_1 = p_2 = 2$, $\Omega_1 = \Omega_2 = \Omega$, $\mathscr{F}_1 = \mathscr{M}_1$, $\mathscr{F}_2 = \mathscr{M}_2$, and $P_1 = P_2 = P$ (restricted to $\mathscr{M}_1$ or $\mathscr{M}_2$), $\mathscr{G}_1 = \mathscr{S}(\mathscr{F}_1)$, and $B = (1/2) \cdot \mathrm{Cov}$ (a product form). Now in the proof of Lemma 3.7(iii) we have that $q = 2$, $\Delta^{(1)} \leqslant \varepsilon/6$, T is defined by $T(f_1) = \frac{1}{2} \cdot [E(f_1 \mid \mathscr{G}_2) - Ef_1]$, and by Claim 0, $P(|T(f_1)| > t) \leqslant [(\varepsilon/t) \cdot \|f_1\|_2]^2$ $\forall\, t > 0$, $\forall\, f_1 \in \mathscr{G}_1$. Also a simple calculation gives $\|T\|_{2 \to 2} = \|B\|_{(2,2)} = \frac{1}{2}\rho(\mathscr{M}_1, \mathscr{M}_2) > \frac{1}{6} \cdot \varepsilon \cdot (1 - \log \varepsilon)^{1/2}$. Since T is a product operator, it is now clear that Theorem 2.1 is sharp up to a constant factor when $n = 2$ and $p_1 = q = 2$.

Acknowledgments

The authors thank M. Rosenblatt for [34], M. Peligrad for a preprint of [28], A. Torchinsky for suggesting the possible connection of this subject with BMO functions, M. Iosifescu for calling the authors' attention to [40], and, for their helpful comments and encouragement, P. Nilsson, J. Bergh, S. Kwapień, G. Bennett, and H. Niemi.

Note added in proof. The authors have constructed examples which show that Theorems 2.1, 3.6, and 4.1(vi) are within a constant factor of being sharp, for every choice of parameters meeting the given specifications. S. Janson has shown that of the measures of dependence mentioned in the last sentence of Remark 4.1, no two are equivalent outside of Theorem 4.1(v) (fourth equation). S. Janson has also shown that the inequality at the beginning of the second line of Theorem 4.2 can be removed; his proof of this more general version of Theorem 4.2 involves some extra work using interpolation techniques.

References

[1] Bergh, J., and Löfström, J. (1976). *Interpolation Spaces.* Springer-Verlag, Berlin.

[2] Blum, J. R., Hanson, D. L., and Koopmans, L. H. (1963). On the strong law of large numbers for a class of stochastic processes. *Z. Wahrsch. Verw. Gebiete* **2** 1–11.

[3] Bradley, R. C. (1983). Equivalent measures of dependence. *J. Multivariate Anal.* **13** 167–176.

[4] Bryc, W. (1983). On almost sure convergence and integrability of sums of weakly dependent random vectors. *Revue Roumanie Math. Pures Appl.* **28** 265–274.

[5] Cogburn, R. (1960). Asymptotic properties of stationary sequences. *Univ. Calif. Publ. Statist.* **3** 99–146.

[6] Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory Probab. Appl.* **13** 691–696.

[7] DAVYDOV, Y. A. (1970). The invariance principle for stationary processes. *Theory Probab. Appl.* **15** 487–498.
[8] DEHLING, H., AND PHILIPP, W. (1982). Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.* **10** 689–701.
[9] DENKER, M., AND KELLER, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete* **64** 505–522.
[10] DEO, C. H. (1973). A note on empirical processes of strong mixing sequences. *Ann. Probab.* **1** 870–875.
[11] DMITROVSKII, V. A., ERMAKOV, S. V., AND OSTROVSKII, E. I. (1983). Central limit theorem for the Banach-valued weakly dependent random variables. *Teor. Veroyatnost. Primenen.* **28** 83–97. [In Russian]
[12] DVORETZKY, A. (1972). Asymptotic normality for sums of dependent random variables. In *Sixth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 513–535.
[13] GEBELEIN, H. (1941). Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung. *Z. Angew. Math. Mech.* **21** 364–379.
[14] GUSTAVSSON, J., AND PEETRE, J. (1977). Interpolation of Orlicz spaces. *Studia Math.* **60** 33–59.
[15] HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.
[16] HIRSCHFELD, H. O. (1935). A connection between correlation and contingency. *Proc. Cambridge Philos. Soc.* **31** 520–524.
[17] IBRAGIMOV, I. A. (1959). Some limit theorems for strictly stationary stochastic processes. *Dokl. Akad. Nauk. SSSR* **125** 711–714. [In Russian]
[18] IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7** 349–382.
[19] IBRAGIMOV, I. A. (1975). A note on the central limit theorem for dependent random variables. *Theory Probab. Appl.* **20** 135–141.
[20] IBRAGIMOV, I. A., AND LINNIK, Y. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen.
[21] IOSIFESCU, M. (1980). Recent advances in mixing sequences of random variables. In *Third International Summer School on Probability Theory and Mathematical Statistics, Varna 1978*, pp. 111–138. Publishing House of the Bulgarian Academy of Sciences, Sofia.
[22] KESTEN, H., AND O'BRIEN, G. L. (1976). Examples of mixing sequences. *Duke Math. J.* **43** 405–415.
[23] LINDENSTRAUSS, J., AND TZAFRIRI, L. (1977). *Classical Banach Spaces* I. Ergebnisse Math. Grenzgebiete. Springer-Verlag, Berlin.
[24] MAL'TSEV, V. V., AND OSTROVSKII, E. I. (1982). Central limit theorem for stationary processes in Hilbert space. *Theory Probab. Appl.* **27** 357–359.
[25] MASE, S. (1982). Properties of fourth-order strong mixing rates and its application to random set theory. *J. Multivariate Anal.* **12** 549–561.
[26] O'NEIL, R. (1963). Convolution operators and $L(p, q)$ spaces. *Duke Math. J.* **30** 129–142.
[27] PEETRE, J. (1970). A new approach in interpolation spaces. *Studia Math.* **34** 23–42.
[28] PELIGRAD, M. (1983). A note on two measures of dependence and mixing sequences. *Adv. Appl. Probab.* **15** 461–464.
[29] PIETSCH, A. (1980). *Operator Ideals.* North–Holland, Amsterdam.
[30] RIETZ, R. E. (1974). A proof of the Grothendieck inequality. *Israel J. Math.* **19** 271–276.
[31] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47.
[32] ROSENBLATT, M. (1970). Density estimates and Markov sequences. In *Nonparametric*

Techniques in Statistical Inference (M. L. Puri, Ed.), pp. 199–210. Cambridge Univ. Press, London.

[33] Rosenblatt, M. (1971). *Markov Processes, Structure and Asymptotic Behavior.* Springer-Verlag, Berlin.

[34] Rosenblatt, M. (1981). Private communication.

[35] Statulevičius, V. (1973). Asymptotic analysis of the distribution of sums of dependent random variables and the regularity conditions. *Abstracts of Comm., Internat. Conference on Probability and Statist., Vilnius, 1973*, pp. 249–252.

[36] Statulevičius, V. (1975). Limit theorems for dependent random variables under various regularity conditions. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1974*, pp. 173–181. Canadian Mathematical Congress.

[37] Stein, E. M., and Weiss, G. (1959). An extension of a theorem of Marcinkiewicz and some of its applications. *J. Math. Mech.* **8** 263–284.

[38] Volkonskii, V. A., and Rozanov, Y. A. (1959). Some limit theorems for random functions, I. *Theory Probab. Appl.* **4** 178–197.

[39] Withers, C. L. (1981). Central limit theorems for dependent random variables, I. *Z. Wahrsch. Verw. Gebiete* **57** 509–534.

[40] Zuev, N. M. (1981). Estimation of covariance of random variables satisfying the Cramer condition and mixing in the sense of Ibragimov. *Vestnik Beloruss. Gos. Univ. Ser. I*, No. 2, 42–44. [In Russian]

[41] Žurbenko, I. G. (1973). Mixing conditions in random processes. In *Random Processes and Statistical Inference, No. 3*, pp. 57–62. Izdat. "Fan" Uzbek. SSR, Tashkent. [In Russian]

[42] Žurbenko, I. G., and Zuev, N. M. (1975). Estimates of the higher spectral densities of stationary processes that satisfy Cramer's condition with mixing in the sense of Rosenblatt. *Litovsk. Mat. Sb.* **15** 111–124. [In Russian]