The Annals of Probability
1993, Vol. 21, No. 3, 1494–1542

LIMIT THEOREMS FOR U-PROCESSES[1]

BY MIGUEL A. ARCONES AND EVARIST GINÉ

Mathematical Sciences Research Institute, Berkeley and University of Connecticut

Necessary and sufficient conditions for the law of large numbers and sufficient conditions for the central limit theorem for U-processes are given. These conditions are in terms of random metric entropies. The CLT and LLN for VC subgraph classes of functions as well as for classes satisfying bracketing conditions follow as consequences of the general results. In particular, Liu's simplicial depth process satisfies both the LLN and the CLT. Among the techniques used, randomization, decoupling inequalities, integrability of Gaussian and Rademacher chaos and exponential inequalities for U-statistics should be mentioned.

1. Introduction. Let $(S, \mathscr{S}, P)$ be a probability space and let $X_i \colon S^{\mathbb{N}} \to S$ be the coordinate functions [$\{X_i\}$ is thus an i.i.d. sequence with $\mathscr{L}(X_i) = P$]. Let $\mathscr{F}$ be a class of measurable real functions on S^m. The U-process based on P and indexed by $\mathscr{F}$ is

$$\begin{aligned} U_m^n(f, P) &:= U_m^n(f) \\ &:= \frac{(n-m)!}{n!} \sum_{(i_2, \dots, i_m) \in I_m^n} f(X_{i_1}, \dots, X_{i_m}), \qquad f \in \mathscr{F}, \end{aligned} \tag{1.1}$$

where $I_m^n = \{(i_1, \dots, i_m) \colon i_j \in \mathbb{N},\ 1 \le i_j \le n;\ i_j \ne i_k \text{ if } j \ne k\}$. These processes appear often in statistics. For instance, Liu's simplicial depth process [Liu (1990)],

$$D_n(x) = \binom{n}{k+1}^{-1} \sum_{1 \le i_1 < \cdots < i_{k+1} \le n} I_{S(X_{i_1}, \dots, X_{i_{k+1}})}(x), \qquad x \in \mathbb{R}^k,$$

where X_i are i.i.d. $\mathbb{R}^k$-valued random variables and $S(x_1, \dots, x_{k+1})$ is the simplex of $\mathbb{R}^k$ determined by $x_1, \dots, x_{k+1} \in \mathbb{R}^k$, is a U-process of order $k+1$ indexed by the class $\mathscr{F} = \{I_{C(x)} \colon x \in \mathbb{R}^k\}$, $C(x) \in \mathbf{S}^{k+1} = (\mathbb{R}^k)^{k+1}$ being the set of all simplices of $\mathbb{R}^k$ that contain x. Nolan and Pollard (1987, 1988) study the law of large numbers and the central limit theorem for U-processes of order $m = 2$ and give also some interesting examples from density estimation and statistics of directions. Their study parallels that of empirical processes: They give sufficient conditions for the central limit theorem to hold in terms of

Received January 1991; revised March 1992.
[1]Partially supported by NSF Grant DMS-90-00132 and PSC-CUMY Grant 661376, carried out at the City University of New York and at the University of Connecticut.
AMS 1991 *subject classifications*. Primary 60F17, 62E20; secondary 60F15, 60B12.
Key words and phrases. U-process, uniform central limit theorem, uniform law of large numbers, metric entropy.

integrals of random entropies, as done for empirical processes in, for example, Section 8 of Giné and Zinn (1984). Of course, they used a symmetrization technique and an exponential inequality adapted to the new situation. The object of this article is to further study the limit theory of U-processes, without restriction to the case $m = 2$. We also follow patterns from empirical process theory, and the additional techniques we use include a decoupling inequality [de la Peña (1992)], exponential inequalities for U-statistics (including a new Bernstein type inequality for degenerate U-statistics) and integrability properties (based on hypercontractivity) of Gaussian and Rademacher chaos [Bonami (1970) and Borell (1979)].

In this section we describe the basics about convergence in law of U-statistics and U-processes, and prove two permanence properties that hold in general (i.e., without extra measurability assumptions) namely, that the "CLT property" is preserved by finite unions and by convex hulls of classes of functions, as in the empirical process case. We thank Professor R. M. Dudley for asking about this and for a discussion on the proofs. Section 2 is devoted to the description of some basic facts to be used later. It contains a new Bernstein type inequality for degenerate U-statistics, which is optimal in a certain sense. Section 3 contains a quite complete study of the law of large numbers for U-processes. We obtain a necessary and sufficient condition for $\|U_m^n - P^m\|_{\mathscr{F}} \to 0$ a.s. under measurability, and apply it to several examples in this section and in Section 6; for instance, the law of large numbers for Banach valued U-statistics is obtained as a corollary. In Section 4 we study the central limit theorem for nondegenerate U-statistics. The results are relatively complete—they cover the important VC-subgraph and bracketing cases, and much more. The more difficult degenerate case is considered in Section 5; some of the results depend on the above mentioned Bernstein type inequality. There are examples in each section, but we collect some special ones in Section 6, particulary the previously mentioned simplicial depth process.

Next we introduce some notation and basic concepts. $\mathscr{F}$ will always denote a collection of real measurable functions on S^m, and the words "real" and "measurable" will usually be omitted. Let G_P be the "Brownian bridge" associated to P, that is, G_P is the centered Gaussian process indexed by $L^2(S, P)$ with covariance

$$EG_P(f)G_P(g) = Pfg - (Pf)(Pg), \qquad f, g \in L^2(S, P). \tag{1.2}$$

Then the finite dimensional distributions of $\{n^{1/2}(U_m^n - P^m(f))\colon f \in \mathscr{F}\}$ converge in law to the corresponding finite dimensional distributions of $\{mG_P \circ P^{m-1}(S_m f)\colon f \in \mathscr{F}\}$, where

$$P^{m-1}f(x) = \int \cdots \int f(x_1, \dots, x_{m-1}, x)\, dP(x_1) \cdots dP(x_{m-1}),$$

and $S_m f(x_1, \dots, x_m) = (m!)^{-1}\Sigma f(x_{i_1}, \dots, x_{i_m})$, the sum extended over the $m!$ permutations $(i_1, \dots, i_m)$ of $\{1, \dots, m\}$. As in empirical process theory, we say that the CLT holds for $\{n^{1/2}(U_m^n - P^m)(f)\colon f \in \mathscr{F}\}$ (or for $\mathscr{F}$) if

$\{G_P \circ P^{m-1} \circ S_m(f)\colon f \in \mathscr{F}\}$ is sample continuous on $(\mathscr{F}, \tau_{P,m})$, with

$$(1.3)\quad \tau_{P,m}^2(f,g) = P\big(P^{m-1} \circ S_m(f-g)\big)^2 - \big(P^m(f-g)\big)^2 \qquad f,g \in \mathscr{F}$$

and if

$$(1.4)\qquad n^{1/2}(U_m^n - P^m) \to_{\mathscr{L}} mG_P \circ P^{m-1} \circ S_m \quad \text{in } l^\infty(\mathscr{F}).$$

Convergence in (1.4) is in the sense of Hoffmann-Jørgensen (1991) [see, e.g., Giné and Zinn (1986)]. Proving (1.4) reduces to the central limit theorem for the empirical process indexed by $\{P^{m-1}f\colon f \in \mathscr{F}\}$ together with convergence in probability to 0 of certain remainder terms (given by Hoeffding's decomposition [Hoeffding (1948)]).

If $P^{m-1} \circ S_m(f) = 0$ for all $f \in \mathscr{F}$, then the limit in (1.4) is 0. This case, that is, the case of degenerate $\mathscr{F}$, is mathematically more appealing because it is genuinely nonlinear (although the nondegenerate case seems to appear more often in applications). Among the classes of degenerate functions, we will only consider classes $\mathscr{F}$ of P-canonical (i.e., completely degenerate) functions f since partial degeneracy reduces to the canonical case together with convergence to 0 of remainder terms via Hoeffding's decomposition. We say that f is P-canonical if $P(S_m f)(x_1, \dots, x_{m-1}, \cdot\,) = 0$ for almost all $x_1, \dots, x_{m-1}$. If f_i are P-canonical and $P^m f_i^2 < \infty$, $i \le k < \infty$, then

$$(1.5)\qquad \big(n^{m/2}U_m^n(f_i)\colon i \le k\big) \to_d \big(K_{P,m}(S_m f_i)\colon i \le k\big),$$

where $K_{P,m}(S_m f)$ is an element of the chaos of order m associated to G_P [Rubin and Vitale (1980), Bretagnolle (1983), Dynkin and Mandelbaum (1984), Gregory (1977) and Serfling (1980) for $m = 2$]. For $\phi \in L^2(S, P)$ with $P\phi = 0$ and $P\phi^2 = \sigma_\phi^2$, let $h^\phi(x_1, \dots, x_m) = \phi(x_1) \cdots \phi(x_m)$ and $K_{P,m}(h^\phi) = (m!)^{-1/2}\sigma_\phi^m H_m(G_P(\phi)/\sigma_\phi)$, where H_m is the Hermite polynomial of degree m and leading coefficient 1. The map $h^\phi \to K_{P,m}(h^\phi)$ extends to a linear isometry $S_m f \to K_{P,m}(f)$ between the subspace of symmetric canonical functions of $L_2(S^m, P^m)$ and $\mathscr{K}_m(G_P)$, the Gaussian chaos space of order m [letting, for each $r \in \mathbb{N}$, $\mathscr{P}_r \subset L_2(\Omega, \Sigma, \Pr)$ be the closure of the linear span of the set of real polynomials of degree r in the variables $G_P(f)$, $f \in L_2(S, \mathscr{S}, P)$, $\mathscr{K}_m(G_P)$ is defined as $\mathscr{P}_m \ominus \mathscr{P}_{m-1}$]. We say that a P-canonical $\mathscr{F}$ satisfies the CLT if the process $\{K_{P,m}(S_m f)\colon f \in \mathscr{F}\}$ has a version with bounded uniformly continuous paths in $(\mathscr{F}, e_{P,m})$, where $e_{P,m}(f,g) = \|f-g\|_{L^2(P^m)}$, and if

$$(1.6)\qquad n^{m/2}U_m^n(f) \to_{\mathscr{L}} K_{P,m} \circ S_m(f) \quad \text{in } l^\infty(\mathscr{F}).$$

The central limit theorems (1.4) and (1.6) reduce to an asymptotic equicontinuity condition as follows. If Y_n, Y are random elements taking values in $l^\infty(\mathscr{F})$, then the law of Y is Radon and $Y_n \to_{\mathscr{L}} Y$ in $l^\infty(\mathscr{F})$ if and only if the finite dimensional distributions of Y_n converge in law to those of Y and there exists a pseudometric ρ on $\mathscr{F}$ such that $(\mathscr{F}, \rho)$ is totally bounded and

$$(1.7)\qquad \lim_{\delta \to 0} \limsup_{n \to \infty} \Pr{}^*\Big\{\sup_{\rho(f,g) \le \delta} |Y_n(f) - Y_n(g)| > \varepsilon\Big\} = 0$$

for all $\varepsilon > 0$. If this is the case, then the process Y admits a version with bounded uniformly ρ-continuous paths and conversely, for any ρ for which there is a version of Y in $C_u(\mathscr{F}, \rho)$, the conditions $(\mathscr{F}, \rho)$ totally bounded and (1.7) are necessary and sufficient for the CLT. This result is due to several authors [see, e.g., Andersen, Giné, Ossiander and Zinn (1988), page 282]. A proof of it in a special case, which readily extends to the general case, can be found in Giné and Zinn [(1986), Theorem 1.1.3]. This criterion for weak convergence will be used throughout in this article. In the CLT for U-statistics with Gaussian limits (1.4), the distance ρ in (1.7) can be taken to be $\tau_{P,m}(f,g) = e_{P,1}(\pi_{1,m} \circ S_m(f), \pi_{1,m} \circ S_m(g))$. Arcones (1991) shows that if the limit law of a degenerate U-statistic is Radon, then it has a version in $C_u(\mathscr{F}, e_{P,m})$ and $(\mathscr{F}, e_{P,m})$ is totally bounded. As a consequence, the CLT for U-processes with canonical kernels (1.6) holds if and only if $C_u(\mathscr{F}, e_{P,m})$ is totally bounded and (1.7) holds with $\rho = e_{P,m}$. [The same applies to the general CLT (1.10) and the distance (1.12).]

Hoeffding's decomposition of a U-statistic will be repeatedly used, so we give it here together with some notation. The operator $\pi_{k,m}^P = \pi_{k,m}$ acts on P^m-integrable functions $h\colon S^m \to \mathbb{R}$ as follows:

$$\pi_{k,m} h(x_1, \dots, x_k) = (\delta_{x_1} - P) \cdots (\delta_{x_k} - P) P^{m-k} h,$$

where $Q_1 \cdots Q_m h = \int \cdots \int h(x_1, \dots, x_m)\, dQ_1(x_1) \cdots dQ_m(x_m)$. Note that $\pi_{k,m} h$ is a P-canonical function of k variables. Hoeffding's decomposition is as follows: For all P^m-integrable functions $f\colon S^m \to \mathbb{R}$,

$$U_m^n(f) = \sum_{k=0}^{m} \binom{m}{k} U_k^n(\pi_{k,m} \circ S_m f). \tag{1.8}$$

The first term is just $P^m f = P^m(U_m^n(f))$.

In the previous paragraphs we have implicitly assumed that the functional $f \to U_m^n(f) - P^m f$ is in $l^\infty(\mathscr{F})$. We will assume, without further mention, a little more, namely,

$$\sup_{f \in \mathscr{F}} |\pi_{k,m} S_m f(x_1, \dots, x_k)| < \infty \quad \text{for all } x_1, \dots, x_k \in S \text{ and } k = 0, 1, \dots, m. \tag{1.9}$$

We must also impose some measurability requirements on the classes $\mathscr{F}$. In fact the classes $\mathscr{F}$ must satisfy measurability conditions allowing for: (1) replace outer probability and outer expectation by probability and expectation, respectively; and (2) randomize (and "unrandomize") by Rademacher or normal multipliers and use Fubini's theorem, both in expressions involving not only $\sup_{f \in \mathscr{F}} |(U_m^n - P^m)(f)|$ but also $\sup_{f \in \mathscr{F}} |U_k^n(\pi_{k,m} S_m f)|$ for all $k \le m$, as well as sups of the same expressions over certain subsets of $\{f - g\colon f, g \in \mathscr{F}\}$, denoted below by $\mathscr{F}_\delta'$. If this is the case, we say that $\mathscr{F}$ is *measurable*. A very general sufficient condition for $\mathscr{F}$ to be measurable is that $\mathscr{F}$ be image admissible Suslin [Dudley (1984), Section 10] and that $\mathscr{F}$ satisfy (1.9). As noted by Dudley, personal communication, if $\mathscr{F}$ satisfies these two conditions then the classes $\pi_k \mathscr{F} := \{\pi_{k,m} S_m f\colon f \in \mathscr{F}\}$ are also image admissible Suslin: By the definition, in order to see this it suffices to observe that if T:

$Y \to \mathscr{F}$ satisfies that $(t, x_1, \ldots, x_m) \to T(t)(x_1, \ldots, x_m)$ is jointly measurable, where $(Y, \mathscr{Y})$ is a measurable space, then the map $(t, x_1, \ldots, x_m) \to \int T(t)(x_1, \ldots, x_m)\, dP(x_1) \cdots dP(x_k)$ is also jointly measurable (by a monotone class argument). Then, Theorem 10.3.2 in Dudley (1984) shows that the operations (1) and (2) on expressions involving the above sups are allowed.

Our notation conforms in general with that of Giné and Zinn (1984). For instance $\|\phi\|_{\mathscr{F}} = \sup\{|\phi(f)|: f \in \mathscr{F}\}$ if $\phi \in l^\infty(\mathscr{F})$, the envelope F of $\mathscr{F}$ is $\sup_{f \in \mathscr{F}} |f|$ and so on. We say that $f: S^m \to \mathbb{R}$ is symmetric if $S_m f = f$.

Next we give two permanence properties of the CLT for U-processes. In this article we only consider the two cases (1.4) (the nondegenerate case) and (1.6) (the P-canonical or completely degenerate case). If the class $\mathscr{F}$ consists of square integrable functions which are degenerate of order $r - 1$ or larger, that is, such that

$$U_m^n(f) - P^m(f) = \sum_{k=r}^{m} \binom{m}{k} U_k^n(\pi_{k,m} \circ S_m f),$$

then, for every $f \in \mathscr{F}$, the sequence $\{n^{r/2}(U_m^n(f) - P^m f)\}$ has the same limit in law as $\{\binom{m}{r} U_r^n(\pi_{r,m} \circ S_m f)\}$, which is $K_{P,r} \circ \pi_{r,m} \circ S_m(f)$. We then say that $\mathscr{F}$ satisfies the CLT if the process $\{K_{P,r} \circ \pi_{r,m} \circ S_m(f): f \in \mathscr{F}\}$ has a version with almost all its trajectories bounded and uniformly continuous for $e_{P,r} \circ \pi_{r,m} \circ S_m$ and if

$$n^{r/2}(U_m^n(f)P - P^m f) \to_{\mathscr{L}} \binom{m}{r} K_{P,r} \circ \pi_{r,m} \circ S_m(f) \quad \text{in } l^\infty(\mathscr{F}), \tag{1.10}$$

(1.4) and (1.6) correspond, respectively, to $r = 1$ and $r = m$ in (1.10). Since for every f, $\pi_{r,m}^P \circ S_m(f)$ is a P-canonical function, a modification of Corollary 4.2 reduces the general case to the P-canonical case. So, without loss of generality we can (under some integrability and measurability conditions) restrict our attention to the P-canonical case. In the following sections we only consider the P-canonical and the nondegenerate cases, but in this section we prove the previously mentioned permanence properties in general.

PROPOSITION 1.1. *If $\mathscr{F}_1$ and $\mathscr{F}_2$ are classes of functions on S^m, degenerate of order at least $r - 1$ for some fixed r, $1 \le r \le m$, and if both satisfy the CLT* (1.10) *for P, then so does the class $\mathscr{F}_1 \cup \mathscr{F}_2$.*

PROOF. Let us denote by K the process $K_{P,r} \circ \pi_{r,m} \circ S_m$ and by $e(f, g)$ the pseudodistance $e_{P,r}(\pi_{r,m} \circ S_m(f), \pi_{r,m} \circ S_m(g))$. Then, K has versions which are chaos random variables with values in $C_u(F_i, e)$, $i = 1, 2$ [Arcones (1991)], and therefore [e.g., Arcones and Giné (1991)] it has an expansion, which we keep denoting by K, as follows:

$$\begin{aligned} K(f) = (m!)^{-1} \sum_{i_1, \ldots, i_m = 1}^{\infty} & E\Big[K(f) \Pi_{j \ge 1} H_{m_j(i_1, \ldots, i_m)}(g_j)\Big] \\ & \times \Pi_{j \ge 1} H_{m_j(i_1, \ldots, i_m)}(g_j), \end{aligned} \tag{1.11}$$

with convergence taking place uniformly a.s., where $m_j(i_1, \dots, i_m) = \sum_{k=1}^m I(i_k = j)$, H_m is the Hermite polynomial of degree m and leading coefficient 1, and $\{g_j\}$ is an ortho-Gaussian sequence. Since the coefficients $E[K(f)\Pi_{j \geq 1} H_{m_j(i_1, \dots, i_m)}(g_j)]$ are uniformly continuous on $(\mathscr{F}_1 \cup \mathscr{F}_2, e)$ it follows that the process $\{K(f): f \in \mathscr{F}_1 \cup \mathscr{F}_2\}$ has almost all of its trajectories in $C_u(\mathscr{F}_1 \cup \mathscr{F}_2, e)$.

As mentioned in comments following (1.7), the CLT for F_i, $i = 1, 2$, implies that for $\varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \Pr{}^*\left\{n^{r/2} \|U_m^n - P^m\|_{\mathscr{F}_i'(\delta, e)} \geq \varepsilon\right\} = 0, \tag{1.12}$$

with $\mathscr{F}_i'(\delta, e) = \{f - g: f, g \in \mathscr{F}_i, e(f, g) \leq \delta\}$ and the pseudometric spaces $(\mathscr{F}_i, e)$, $i = 1, 2$, are totally bounded. For each $\delta > 0$, let τ_δ: $\mathscr{F}_1 \cup \mathscr{F}_2 \to \mathscr{F}_1 \cup \mathscr{F}_2$ be a map with finite range, with $\tau_\delta(f) \in \mathscr{F}_i$ for $f \in \mathscr{F}_i$, $i = 1, 2$, and such that $e(\tau_\delta(f), f) \leq \delta$ for all $f \in \mathscr{F}_1 \cup \mathscr{F}_2$. (Such a map exists by total boundedness.) Then, the following inequalities follow by (1.12), finite dimensional convergence in distribution of U-statistics and sample continuity of K on $\mathscr{F}_1 \cup \mathscr{F}_2$:

$$\begin{aligned}
&\lim_{\delta \to 0} \limsup_{n \to \infty} \Pr{}^*\left\{n^{r/2} \|U_m^n - P^m\|_{(\mathscr{F}_1 \cup \mathscr{F}_2)'(\delta, e)} \geq 3\varepsilon\right\} \\
&\quad \leq \lim_{\delta \to 0} \limsup_{n \to \infty} \sum_{i=1}^{2} \Pr{}^*\left\{n^{r/2} \|U_m^n - P^m\|_{\mathscr{F}_i'(\delta, e)} \geq \varepsilon\right\} \\
&\qquad + \lim_{\delta \to 0} \limsup_{n \to \infty} \Pr{}^*\Bigg\{\max_{f, g \in \mathscr{F}_1 \cup \mathscr{F}_2;\, e(\tau_\delta(f), \tau_\delta(g)) \leq 3\delta} n^{r/2} \\
&\qquad\qquad \times \left|(U_m^n - P^m)(\tau_\delta(f) - \tau_\delta(g))\right| \geq \varepsilon\Bigg\} \\
&\quad \leq \lim_{\delta \to 0} \Pr{}^*\left\{\|K\|_{(\mathscr{F}_1 \cup \mathscr{F}_2)'(3\delta, e)} \geq \varepsilon\right\} = 0.
\end{aligned}$$

Therefore, the class $\mathscr{F}_1 \cup \mathscr{F}_2$ satisfies the CLT. □

PROPOSITION 1.2. *Let $\mathscr{F}$ be a class of functions on S^m, degenerate of order at least r, $1 \leq r \leq m$, satisfying the CLT* (1.10). *Let $\mathscr{H}$ be the symmetric convex hull of $\mathscr{F}$ and let $\mathscr{G}$ be the set of all limits of functions in $\mathscr{H}$, simultaneously pointwise and in $\mathscr{L}_2(P^m)$. Then, $\mathscr{G}$ also satisfies the CLT* (1.10).

PROOF. This follows by the a.s. representation theorem in Dudley [(1985), Theorem 4.1], the linearity of $U_m^n - P^m$ and of K [cf. (1.11)], and their continuity for the simultaneous convergence. □

Finally we should mention that, although we restrict our attention to U-processes of the form (1.1), the results of this article extend to more general situations, such as multisample U-processes.

2. Preliminaries. Randomization by Rademacher variables plays a role in U-processes similar to the role it plays for regular empirical processes due to the fact that U-processes can be "decoupled." We state here the pertinent result from de la Peña (1992) [see also Kwapień (1987) for other decoupling results and related techniques]. The statement of the next proposition involves several classes of functions $\mathscr{F}_{i_1,\dots,i_m}$; to ease notation $\|\Sigma f_{i_1,\dots,i_m}\|_{\mathscr{F}}$ will denote $\sup_{f_{i_1,\dots,i_m} \in \mathscr{F}_{i_1,\dots,i_m}} |\Sigma f_{i_1,\dots,i_m}|$.

PROPOSITION 2.1 [de la Peña (1992)]. *Let $\mathscr{F}_{i_1,\dots,i_m}$, $(i_1,\dots,i_m) \in I_m^n$, $m \le n < \infty$, be classes of functions in $L_1(P^m)$ and let $\phi\colon [0,\infty) \to \mathbf{R}$ be a convex, increasing function. Let $\{X_i\colon i \in \mathbf{N}\}$ be i.i.d. and let $\{X_i^{(k)}\colon i \in \mathbf{N}\}_{k=1}^m$ be i.i.d. copies of $\{X_i\colon i \in \mathbf{N}\}$. Then*

$$
\begin{aligned}
&E^*\phi\left(\left\| \sum_{(i_1,\dots,i_m)\in I_m^n} f_{i_1,\dots,i_m}(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{F}}\right) \\
&\qquad \lesssim E^*\phi\left(\left\| \sum_{(i_1,\dots,i_m)\in I_m^n} f_{i_1,\dots,i_m}(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}) \right\|_{\mathscr{F}}\right).
\end{aligned}
\tag{2.1}
$$

If moreover the functions f satisfy $f_{i_1,\dots,i_m} = f_{\sigma(i_1),\dots,\sigma(i_m)}$ for each permutation σ and each $i_1,\dots,i_m$, and they are symmetric (i.e., the classes are symmetric and the functions in each class are themselves symmetric), then the reverse inequality holds. If the classes $\mathscr{F}_{i_1,\dots,i_m}$ are measurable and consist only of P-canonical functions, then the right-hand side of (2.1) *is equivalent to*

$$
E\phi\left(\left\| \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_m}^{(m)} f_{i_1,\dots,i_m}(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}) \right\|_{\mathscr{F}}\right).
\tag{2.2}
$$

Under the symmetry condition, by further use of (2.1), *this is also equivalent to*

$$
E\phi\left(\left\| \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1} \cdots \varepsilon_{i_m} f_{i_1,\dots,i_m}(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{F}}\right),
\tag{2.3}
$$

where $\{\varepsilon_i\}$ are i.i.d. independent of $\{X_i\}$ and $\{\varepsilon_i^{(k)}\}$ are independent copies of $\{\varepsilon_i\}$, independent of $\{X_i^{(k)}\colon i \in \mathbf{N}\}_{k=1}^m$.

"Equivalent" for $E\phi(A)$ and $E\phi(B)$ means that there are constants $c_i = c_i(m)$ (independent of $\phi, n, \mathscr{F}$) such that $E\phi(c_1 A) \le E\phi(B) \le E\phi(c_2 A)$ and $\lesssim$ means that $E\phi(c_1 A) \le E\phi(B)$.

The equivalence between (2.1) and (2.3) allows us the use of Khinchin-type inequalities for the Rademacher chaos.

PROPOSITION 2.2 [Borell (1979)]. *Let $x_{i_1,\dots,i_m}$ be elements in a Banach space $(B, \|\cdot\|)$ and let $\{\varepsilon_i\}$ be as above. Then, letting $X = \sum_{i_1<\cdots<i_m\le n} \varepsilon_{i_1}\cdots\varepsilon_{i_m} x_{i_1,\dots,i_m}$,*

$$
(E\|X\|^p)^{1/p} \le \left(\frac{p-1}{q-1}\right)^{m/2} (E\|X\|^q)^{1/q} \quad \textit{for all } 1 < q < p < \infty.
$$

Proposition 2.2 will play the role of the exponential inequalities for sub-Gaussian processes in empirical processes theory. Another very useful exponential inequality is Bernstein's (or Prohorov's) for sums of independent random variables. This inequality has a well known extension to U-statistics [Hoeffding (1963)] which will be useful for the treatment of the nondegenerate case. We present below a new Bernstein type inequality for degenerate U-statistics.

PROPOSITION 2.3. *Let* $\|f\|_\infty \le c$, $Ef(X_1, \dots, X_m) = 0$ *and* $\sigma^2 = Ef^2(X_1, \dots, X_m)$. *Then for any* $t > 0$:

(a) $$P\{U_m^n(f, P) > t\} \le \exp\left\{-\frac{[n/m]t^2}{2\sigma^2 + (2/3)ct}\right\} \qquad \text{[Hoeffding (1963)]}.$$

(b) $$P\{U_m^n(f, P) > t\} \le \exp\{-[n/m]t^2/c^2\} \qquad \text{[Hoeffding (1963)]}.$$

If moreover f is P-canonical there are constants c_i depending only on m such that

(c) $$P\left\{\left|n^{-m/2}\sum_{(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m})\right| \ge t\right\} \le c_1\exp\left\{-\frac{c_2 t^{2/m}}{\sigma^{2/m} + \left(ct^{1/m}n^{-1/2}\right)^{2/(m+1)}}\right\}$$

and

(d) $$P\left\{\left|n^{-m/2}\sum_{(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m})\right| \ge t\right\} \le c_1\exp\left\{-c_2(t/c)^{2/m}\right\}.$$

These inequalities also hold for decoupled U-statistics.

PROOF OF PROPOSITION 2.3(c). We may assume that f is symmetric. We first note that, by Proposition 2.2, if $X = \sum_{i_1<\dots<i_m\le n}\varepsilon_{i_1}\cdots\varepsilon_{i_m}a_{i_1\cdots i_m}$ with $s^2 = \sum_{i_1<\dots<i_m\le n}a^2_{i_1\cdots i_m}$, then there exist $c_1(\alpha, m), c_2(\alpha, m) \in (0,\infty)$ such that for all $t > 0$ and $0 < \alpha < 2/m$,

$$Ee^{t|X|^\alpha} \le c_1\exp\left\{c_2(s^\alpha t)^{1/(1-\alpha m/2)}\right\}. \tag{2.4}$$

[Proof of inequality (2.4): Developing $\exp\{\lambda|X/s|^{2/m}\}$ and applying Proposition 2.2 gives that if $0 < \lambda m/2e$ then there is $M(\lambda) < \infty$ independent of s such that $E\exp\{\lambda|X/s|^{2/m}\} < M(\lambda)$. To relate $|X|^{2/m}$ to $|X|^\alpha$, $\alpha < 2/m$, we use the inequality $|ab| \le 1/p|a|^p + 1/q|b|^q$, $1/p + 1/q = 1$, $1 < p < \infty$, $a, b \in \mathbb{R}$, which gives $t|X|^\alpha \le \alpha m/2|cX/s|^{2/m} + (1 - \alpha m/2)(s^\alpha t/c^\alpha)^{1/1-\alpha m/2}$ for any $c > 0$. Taking c so that $\lambda := \alpha m c^{2/m}/2 < m/2e$ and applying the previous inequality yields (2.4).] Now we take α so that the exponent of s in (2.4) is 2, that is, $\alpha = 2/(m+1)$, apply (2.3) in Proposition 2.1 to the class of functions $\{t^{1/\alpha}f\colon f \in \mathscr{F}\}$ for a convex function Ψ satisfying $\delta\Psi(x) \le \exp x^\alpha \le \Psi(x)$ for all $x > 0$ and some $\delta > 0$, and then (2.4) with $a_{i_1,\dots,i_m} = f(X_{i_1},\dots,X_{i_m})$, to obtain

$$
\begin{aligned}
(2.5)\qquad & E\exp\left\{t\left|n^{-m/2}\sum_{i_1<\cdots<i_m\le n} f(X_{i_1},\dots,X_{i_m})\right|^{2/(m+1)}\right\}\\
&\le c_1 E\exp\left\{c_2 t^{m+1} n^{-m}\sum_{i_1<\cdots<i_m\le n} f^2(X_{i_1},\dots,X_{i_m})\right\}\\
&\le c_1\exp\{c_2\sigma^2 t^{m+1}\}\\
&\quad\times E\exp\left\{c_2 t^{m+1} n^{-m}\sum_{i_1<\cdots<i_m\le n}\left(f^2(X_{i_1},\dots,X_{i_m})\right.\right.\\
&\qquad\qquad\left.\left.-Ef^2(X_1,\dots,X_m)\right)\right\}.
\end{aligned}
$$

The constants c_1, c_2 depend only on m, but may not be the same as in (2.4). Since $(m!(n-m)!/n!)\sum_{i_1<\cdots<i_m\le n} f(X_{i_1},\dots,X_{i_m})$ is the average of $W(X_{i_1},\dots,X_{i_n})$ over all the permutations $(i_1,\dots,i_n)$ of $1,\dots,n$ with $W(X_1,\dots,X_n) = k^{-1}\sum_{j=0}^{k-1} f(X_{jm+1},\dots,X_{(j+1)m})$ and $k = [n/m]$ [Hoeffding (1963) e.g., Serfling (1980)], we have, by convexity,

$$
\begin{aligned}
(2.6)\qquad & E\exp\left\{c_2 t^{m+1} n^{-m}\sum_{i_1<\cdots<i_m\le n}\left(f^2(X_{i_1},\dots,X_{i_m})\right.\right.\\
&\qquad\qquad\left.\left.-Ef^2(X_1,\dots,X_m)\right)\right\}\\
&\le E\exp\left\{2mc_2 t^{m+1} n^{-1}\sum_{i=1}^{[n/m]}\left(f^2(X_{1+(i-1)m},\dots,X_{im})\right.\right.\\
&\qquad\qquad\left.\left.-Ef^2(X_1,\dots,X_m)\right)\right\}.
\end{aligned}
$$

A form of Bernstein's inequality for i.i.d. random variables is

$$
E\exp\left\{tn^{-1/2}\sum_{i=1}^{n}\xi_i\right\} \le \exp\left\{\frac{t^2\sigma^2}{2-(2/3)tcn^{-1/2}}\right\},\ |t| < \frac{3n^{1/2}}{c},
$$

where ξ are i.i.d., $|\xi| \le c$, $E\xi^2 = \sigma^2$, $E\xi = 0$. By applying this inequality in (2.6) and then in (2.5), we obtain

$$
\begin{aligned}
(2.7)\qquad & E\exp\left\{t\left|n^{-m/2}\sum_{i_1<\cdots<i_m\le n} f(X_{i_1},\dots,X_{i_m})\right|^{2/(m+1)}\right\}\\
&\le c_1\exp\left\{c_2 t^{m+1}\sigma^2 + \frac{4mc_2^2 t^{2(m+1)}\sigma^2 c^2}{2n-(8/3)mc_2 c^2 t^{m+1}}\right\}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&P\left\{n^{-m/2}\left|\sum_{i_1<\cdots<i_m\le n} f(X_{i_1},\dots,X_{i_m})\right|\ge u\right\}\\
(2.8)\qquad &\le c_1\exp\left\{-tu^{2/(m+1)}+c_2t^{m+1}\sigma^2+\frac{4mc_2t^{2(m+1)}\sigma^2c^2}{2n-(8/3)mc^2t^{m+1}}\right\}.
\end{aligned}
$$

If we take

$$t=\left(u^{2/(m+1)}/c_2(m+1)\sigma^2\right)^{1/m}$$

[obtained by minimizing $-tu^{2/(m+1)}+c_2t^{m+1}\sigma^2$], the sum of the first two terms in the exponent are of the order $-c_3(u/\sigma)^{2/m}$. This will be the order of the whole exponent if the third term is smaller than a small constant times $(u/\sigma)^{2/m}$. Therefore, there are K and $c_2'<\infty$ such that, under the condition

$$\frac{u^{2/m}c^2}{n\sigma^{2(m+1)/m}}\le K, \tag{2.9}$$

we have $P\{n^{-m/2}|\sum_{(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m})|\ge u\}\le c_1\exp\{-c_2'(u/\sigma)^{2/m}\}$. If condition (2.9) is not satisfied, then $t^{m+1}\cong nc^{-2}$ gives a bound for the right-hand side of (2.8) of the order $c_1\exp\{-c_2''(nu^2/c^2)^{1/(m+1)}\}$. This proves part (c). □

PROOF OF PROPOSITION 2.3(d). As in (2.5),

$$
\begin{aligned}
&E\exp\left\{t\left|n^{-m/2}\sum_{i_1<\cdots<i_m\le n} f(X_{i_1},\dots,X_{i_m})\right|^{2/(m+1)}\right\}\\
&\quad\le c_1\exp\left\{c_2t^{m+1}n^{-m}\sum_{i_1<\cdots<i_m\le n} f^2(X_{i_1},\dots,X_{i_m})\right\}\le c_1\exp\{c_2t^{m+1}c^2\}.
\end{aligned}
$$

Thus $P\{n^{-m/2}|\sum_{i_1<\cdots<i_m\le n} f(X_{i_1},\dots,X_{i_m})|>u\}\le c_1\exp\{-tu^{2/(m+1)}+c_2t^{m+1}c^2\}$ and (d) follows by minimizing the exponent with respect to t. □

The proof for decoupled U-statistics is very similar and is omitted. Note that the inequality in (c) is just Bernstein's when $m=1$. Inequality (c) shows that the tail of the normalized U-statistic $n^{m/2}U_m^n(f)$ for f P-canonical with $P^mf^2=\sigma^2$, $\|f\|_\infty\le c$, is of the order of the tail of the limiting chaos process, namely $\exp\{-c_2(t/m)^{2/m}\}$, only for $t\lesssim\sigma^{m+1}n^{m/2}/c^m$. We show next that this is the correct "breakpoint." If $\{X_i\},\{Y_i\}$ are two independent sequences, each i.i.d. but possibly with different laws, then the expression $(n(n-1))^{-1}\sum_{(i,j)\in I_2^n}X_iY_j$ is a U-statistic [with $h((x,y),(u,v))=(xv+yu)/2$ and $P=\mathscr{L}(X,Y)$]. Let, for n fixed, $\mathscr{L}(V_i)=(1/n)\delta_1+((n-1)/n)\delta_0$, $X_i=V_i-EV_i$ and $\mathscr{L}(Y_i)=(1/2)(\delta_{-1}+\delta_1)$. We show that there exists $M<\infty$ such that the inequality $\Pr\{n^{-1}|\sum_{(i,j)\in I_2^n}X_iY_j|>t\}\le c_1\exp(-c_2t/\sigma)$ cannot happen for $t\ge M\sigma^3n/c^2$ if n is large, where $\sigma^2=n^{-1}$ and $c=1$. Let $a_0>0$ be such that

$$c_1 \exp\left\{-2c_2 a^{3/2}(\log a)^{1/2}\right\} < \Pr\{|\lambda - 1| \geq a\} \Pr\left\{|g| \geq (a \log a)^{1/2}\right\}$$

for all $a \geq a_0$, where λ is Poisson with parameter 1 and g is $N(0,1)$; this is satisfied for all a large enough because this product is of the order of $\exp(-(3/2)a \log a)$ for a large. Let $M = 2a_0^{3/2}(\log a_0)^{1/2}$ and $t = 2a^{3/2}(\log a)^{1/2}n^{-1/2}$ for $a \geq a_0$. Then

$$\Pr\left\{|n^{-1}\Sigma_{I_2^n} X_i Y_j| \geq t\right\} \leq \Pr\{|\Sigma_{i=1}^n X_i| \geq a\}\Pr\left\{|\Sigma_{i=1}^n Y_i| \geq (a \log a)^{1/2} n^{1/2}\right\}$$
$$- \Pr\left\{|\Sigma_{i=1}^n Y_i| \geq a^{3/2}(\log a)^{1/2} n^{1/2}\right\}.$$

This last expression tends to $\Pr\{|\lambda - 1| \geq a\}\Pr\{|g| \geq (a \log a)^{1/2}\}$ as $n \to \infty$. A similar example gives the optimality of the breakpoint for higher order U-statistics (one takes the product of $m - 1$ Rademacher variables and one Bernoulli $1/n$). We thank M. Talagrand for comments regarding the optimality property of the inequality in 2.3(c).

The breakpoint in Bernstein's inequality for degenerate V-statistics satisfying $|h(X_1, \dots, X_m)| \leq \Pi_{i=1}^m g(X_i)$ is different from the above [Borisov (1990)].

Next we state (and indicate the proof of) Hoeffding's extension of Chernoff's inequality for binomial probabilities [see, e.g., Proposition 2.2.5 in Dudley (1984)], which is needed below.

Proposition 2.4 [Hoeffding (1963)]. *Let $\{X_i\}_{i=1}^\infty$ be i.i.d. r.v.'s and let h: $\mathbf{R}^m \to \{0,1\}$ with $Eh = p$. Then, for $p < t < 1$,*

$$P\{U_m^n(h) \geq t\} \leq \left(\frac{p}{t}\right)^{t[n/m]} \left(\frac{1-p}{1-t}\right)^{(1-t)[n/m]}.$$

Proof. For $\lambda > 0$, using the argument preceding (2.6),

$$P\{U_m^n(h) \geq t\} \leq Ee^{\lambda[n/m](U_m^n(h)-t)}$$
$$\leq E \exp \lambda \sum_{j=1}^{[n/m]} \left(h(X_{1+(j-1)m}, \dots, X_{m+(j-1)m}) - t\right)$$
$$= \left(pe^{\lambda(1-t)} + (1-p)e^{-\lambda t}\right)^{[n/m]}.$$

Taking $\lambda = \log((1-p)t/p(1-t))$ gives the bound. □

We will require a now well known metric entropy bound for processes satisfying some regularity. This goes back at least to Dudley (1967); the version here can be found in Fernique (1983) and in Pisier (1983). A Young function Ψ is an increasing convex function with $\Psi(0) = 0$.

Proposition 2.5. *Let (T, d) be a a pseudometric space, let Ψ be a Young function and let $\{X_t, t \in T\}$ be a stochastic process with values in a Banach* *space such that* $E\Psi(\|X_t - X_s\|/d(s,t)) \le 1$ *for all* s,t *with* $d(s,t) < \infty$. *Then*

$$E \sup_{s,t\in T} \|X_t - X_s\| \le 8\int_0^D \Psi^{-1}(N(T,d,\varepsilon))\, d\varepsilon,$$

where $N(T,d,\varepsilon) = \min\{n\colon \exists$ *a covering of* T *by* n *balls of radius* $\le \varepsilon\}$ *and* D *is the diameter of* (T,d).

A basic elementary inequality in the proof of this proposition is that

$$(2.10)\qquad \begin{gathered} E\Psi(\|X_i/c_i\|) \le \alpha,\ i = 1,\dots,N, \\ \text{implies } E\max_{i\le N}\|X_i\| \le \Psi^{-1}(\alpha N)\max_{i\le N}|c_i| \end{gathered}$$

for any set of Banach space valued random variables X_i. (The proof uses only convexity.)

Proposition 2.5 can be strengthened to a bound involving majorizing measures [see Ledoux and Talagrand (1991), Chapter 11] and this would produce somewhat sharper results in what follows but, for simplicity, majorizing measures will not be considered in this paper.

If a process $\{X_t\colon t\in T\}$ satisfies

$$(2.11)\quad \left(E\|X_t - X_s\|^p\right)^{1/p} \le \left(\frac{p-1}{q-1}\right)^{m/2}\left(E\|X_t - X_s\|^q\right)^{1/q}, \qquad 1 < q < p < \infty$$

for some $m \ge 1$ (see Proposition 2.2), and if

$$(2.12)\qquad \rho(s,t) = \left(E\|X_t - X_s\|^2\right)^{1/2},$$

then it follows easily that $E\exp\{(\|X_t - X_s\|/c\rho(s,t))^{2/m}\} \le c$, where $c = c(m) < \infty$ depends only on m. We can apply Proposition 2.5 to $\{X_t\}$ with a Young function of the same order as $\exp(|x|^{2/m})$ at ∞, for instance with $\Psi(x) = \sum_{r=1}^{\infty} x^{2r}/(mr)!$, to obtain the following proposition.

PROPOSITION 2.6. *If* $\{X_t\colon t\in T\}$ *satisfies* (2.11) *and* ρ *is as in* (2.12), *there is a constant* $K = K(m) < \infty$ *such that*

$$(2.13)\qquad E\sup_{s,t\in T}\|X_t - X_s\| \le K\int_0^D [\log N(T,\rho,\varepsilon)]^{m/2}\, d\varepsilon,$$

D *being the* ρ*-diameter of* T. *Moreover, if* T *is finite, that is,* $T = \{1,\dots,N\}$, *and* $N \ge 2$, *then*

$$(2.14)\qquad E\max_{i\le N}\|X_i\| \le K(\log N)^{m/2}\max_{i\le N}\left(E\|X_i\|^2\right)^{1/2}.$$

We will apply Proposition 2.6 not only to Rademacher chaos processes but also to the limiting processes of degenerate U-statistics, namely the Gaussian chaos processes $\{K_{P,m}(f)\colon f\in\mathscr{F}\}$ described in the introduction. In fact, Nelson [(1973), Theorem 3], showed that:

PROPOSITION 2.7.

$$(2.15)\qquad \begin{aligned}&\left(E|K_{P,m}(f)-K_{P,m}(g)|^p\right)^{1/p}\\&\quad\le\left(\frac{p-1}{q-1}\right)^{m/2}\left(E|K_{P,m}(f)-K_{P,m}(g)|^q\right)^{1/q}\end{aligned}$$

for $1<q<p<\infty$. *Therefore Proposition* 2.6 *applies to* $\{K_{P,m}(f)\colon f\in\mathscr{F}\}$ *with* $\rho(f,g)=e_{P,m}(f,g)$.

The following Hoffmann-Jørgensen type inequality will help us treat integrability in some cases.

PROPOSITION 2.8 [Giné and Zinn (1992b)]. *Let* $\mathscr{F}$ *be a measurable class of real functions on* S^m. *There exist finite constants* $c_1(p)$, $c_2(p)$ *and* $c_3(p)\in(0,1)$ *such that*

$$\begin{aligned}&E\left\|\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1}^{(1)}\cdots\varepsilon_{i_m}^{(m)}f\left(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\right)\right\|_{\mathscr{F}}^p\\&\quad\le c_1t_0^p+c_2E\max_{i_m\le n}\left\|\sum_{(i_1,\dots,i_{m-1})\colon(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1}^{(1)}\cdots\varepsilon_{i_{m-1}}^{(m-1)}\right.\\&\qquad\qquad\left.\times f\left(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\right)\right\|_{\mathscr{F}}^p,\end{aligned}$$

where t_0 *is any number satisfying*

$$\Pr{}^*\left\{\left\|\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1}^{(1)}\cdots\varepsilon_{i_m}^{(m)}f\left(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\right)\right\|_{\mathscr{F}}>t_0\right\}\le c_3.$$

If moreover the class $\mathscr{F}$ *consists of P-canonical functions, then the same inequality holds for the U-process* $\{U_m^n(f,P)\colon f\in\mathscr{F}\}$, *possibly with different constants.*

Note that, for $m=2$ and $\mathscr{F}$ uniformly bounded, Proposition 2.8 implies the following: If $\{\|n^{-1}\sum_{(i,j)\in I_2^n}\varepsilon_i^{(1)}\varepsilon_j^{(2)}f(X_i^{(1)},X_j^{(2)})\|_{\mathscr{F}}\}$ is stochastically bounded, then all the powers of this sequence are uniformly integrable.

3. The law of large numbers for *U*-processes. We will prove an analogue of the laws of large numbers for empirical processes of Vapnik and Červonenkis (1981) and Giné and Zinn (1984). We introduce some random entropies that will be used throughout. Given a pseudometric space $(\mathscr{F},e)$ the

ε-covering number of $(\mathscr{F}, e)$ is

$$N(\varepsilon, \mathscr{F}, e) = \min\left\{n: \exists\, f_1, \dots, f_n \in \mathscr{F} \text{ s.t. } \sup_{f \in \mathscr{F}} \min_{i \le n} e(f, f_i) \le \varepsilon\right\}.$$

If $\{X_i\}_{i=1}^{\infty}$ are i.i.d. and $\mathscr{F}$ is a class of real functions on S^m, then we define $N_{n,p}(\varepsilon, \mathscr{F})$ as the (random) ε-covering numbers of $(\mathscr{F}, e_{n,p})$, where $e_{n,p}(f, g) = (U_m^n(|f-g|^p))^{(1/p)\wedge 1}$. If $\{X_i^{(j)}\}_{i=1}^{\infty}$, $j \ge 1$, are i.i.d. copies of $\{X_i\}_{i=1}^{\infty}$, then $N_{n,p}^{\text{dec}}(\varepsilon, \mathscr{F})$ is the (random) ε-covering number of $(\mathscr{F}, e_{n,p}^{\text{dec}})$, where

$$e_{n,p}^{\text{dec}}(f, g) = \left(((n-m)!/n!) \sum_{(i_1, \dots, i_m) \in I_m^n} |(f-g)\big(X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)}\big)|^p\right)^{(1/p)\wedge 1}.$$

We also define some other distances, namely,

$$\tilde{e}_{n,p}(f, g) = \left(\frac{1}{n} \sum_{i_1=1}^{n} \left| \frac{(n-m)!}{(n-1)!} \sum_{(i_2, \dots, i_m):\, (i_1, \dots, i_m) \in I_m^n} (f-g) \times (X_{i_1}, \dots, X_{i_m}) \right|^p\right)^{(1/p)\wedge 1}$$

for $0 < p < \infty$, and

$$\tilde{e}_{n,\infty}(f, g) = \max_{i_1 \le n} \left| \frac{(n-m)!}{(n-1)!} \sum_{(i_2, \dots, i_m):\, (i_1, \dots, i_m) \in I_m^n} (f-g)\big(X_{i_1}, \dots, X_{i_m}\big) \right|$$

along with their associated covering numbers $\tilde{N}_{n,p}(\varepsilon, \mathscr{F})$, as well as the same distances for the decoupled statistic , $\tilde{e}_{n,p}^{\text{dec}}(f, g)$ and $\tilde{N}_{n,p}^{\text{dec}}(\varepsilon, \mathscr{F})$. For $p = 2$,

$$\tilde{e}_{n,2}(f, g) = n^{1/2} \frac{(n-m)!}{(n)!} \left[E_{\varepsilon}\left(\sum_{(i_1, \dots, i_m) \in I_m^n} \varepsilon_{i_1}(f-g)\big(X_{i_1}, \dots, X_{i_m}\big)\right)^2\right]^{1/2}. \tag{3.1}$$

The following is the main result of this section.

THEOREM 3.1. *Let $\mathscr{F}$ be a measurable class of symmetric functions on S^m and let P be a probability measure on $(S, \mathscr{S})$ such that $P^m F < \infty$. Then the following conditions are equivalent*:

(i) $\|U_m^n - P^m\|_{\mathscr{F}} \to 0$ *a.s.*

(ii) $E\|n^{-m} \sum_{(i_1, \dots, i_m) \in I_m^n} \varepsilon_{i_1} f(X_{i_1}, \dots, X_{i_m})\|_{\mathscr{F}} \to 0$.

(iii) $n^{-1} \log \tilde{N}_{n,1}(\varepsilon, \mathscr{F}) \to 0$ *in probability* for all* $\varepsilon > 0$.

PROOF. Since $\|U_m^n - P^m\|_{\mathscr{F}}$ is a reverse submartingale [e.g., Nolan and Pollard (1987)], (i) is equivalent to $E\|U_m^n - P^m\|_{\mathscr{F}} \to 0$ by Doob's reversed submartingale limit theorem [e.g., Dudley (1989), Theorem 10.6.4]. We first show (ii) $\Rightarrow$ (i). In the inequalities that follow c will denote a constant that may vary from line to line. We have

$$
\begin{aligned}
&E\|U_m^n - P^m\|_{\mathscr{F}} \\
&\le c\frac{(n-m)!}{n!}E\bigg\|\sum_{(i_1,\dots,i_m)\in I_m^n}\Big(f\big(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\big) - P^m f\Big)\bigg\|_{\mathscr{F}} \\
&\qquad\qquad \text{by Proposition 2.1} \\
&\le c\frac{(n-m)!}{n!} \\
&\quad \times E\bigg\|\sum_{(i_1,\dots,i_m)\in I_m^n}\Big(f\big(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\big) - f\big(Y_{i_1}^{(1)},\dots,Y_{i_m}^{(m)}\big)\Big)\bigg\|_{\mathscr{F}} \\
&\qquad\qquad \text{where } \big\{Y_i^{(j)}\big\}_{i=1}^{\infty} \text{ is an independent copy of } \big\{X_i^{(j)}\big\}_{i=1}^{\infty} \\
&\le c\sum_{j=1}^{m}\frac{(n-m)!}{n!} \\
&\quad \times E\bigg\|\sum_{(i_1,\dots,i_m)\in I_m^n}\Big(f\big(X_{i_1}^{(1)},\dots,X_{i_j}^{(j)},Y_{i_{j+1}}^{(j+1)},\dots,Y_{i_m}^{(m)}\big) \\
&\qquad\qquad -f\big(X_{i_1}^{(1)},\dots,X_{i_{j-1}}^{(j-1)},Y_{i_j}^{(j)},\dots,Y_{i_m}^{(m)}\big)\Big)\bigg\|_{\mathscr{F}} \\
&\qquad\qquad \text{by the triangle inequality, after replacing the } X\text{'s by the } Y\text{'s, one entry at a time} \\
&\le c\frac{(n-m)!}{n!}E\bigg\|\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1} f\big(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\big)\bigg\|_{\mathscr{F}} \\
&\qquad\qquad \text{where } \{\varepsilon_i\} \text{ is a Rademacher sequence independent of } \big\{X_i^{(j)}\big\} \\
&\le c\frac{(n-m)!}{n!}E\bigg\|\sum_{(i_1,\dots,i_m)\in I_m^n}\Big(\varepsilon_{i_1}^{(1)} f\big(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\big) \\
&\qquad\qquad + \cdots + \varepsilon_{i_m}^{(m)} f\big(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\big)\Big)\bigg\|_{\mathscr{F}} \\
&\qquad\qquad \text{by Jensen's inequality on the } \varepsilon_i^{(j)}\text{'s} \\
&\le c\frac{(n-m)!}{n!}E\bigg\|\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1} f(X_{i_1},\dots,X_{i_m}) \\
&\qquad\qquad + \cdots + \varepsilon_{i_m} f\big(X_{i_1},\dots,X_{i_m}\big)\bigg\|_{\mathscr{F}} \\
&\qquad\qquad \text{by Proposition 2.1 applied to the functions } (y_1+\cdots+y_m)f(x_1,\dots,x_m)\text{, which are symmetric in the variables } (x_i,y_i),\ i\le m \\
&\le c\frac{(n-m)!}{n!}E\bigg\|\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1} f\big(X_{i_1},\dots,X_{i_m}\big)\bigg\|_{\mathscr{F}}.
\end{aligned}
$$

So, (ii) $\Rightarrow$ (i). Now we show (i) $\Rightarrow$ (ii). For simplicity we will consider only the case $m = 2$. For ease of notation we will write, in what follows, $Pf(X_i)$ instead of $\int f(X_i, x)\,dP(x)$, and $Pf(X_j)$ instead of $\int f(x, X_j)\,dP(x)$. We have

$$\begin{aligned}
E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i f(X_i, X_j)\Bigg\|_{\mathscr{F}}
&\le E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i P^2 f\Bigg\|_{\mathscr{F}} + E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i Pf(X_i)\Bigg\|_{\mathscr{F}} \\
&\quad + E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i Pf(X_j)\Bigg\|_{\mathscr{F}} + E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i \pi_{2,2}^P f(X_i, X_j)\Bigg\|_{\mathscr{F}} \\
&\le n^{-1/2}\|P^2 f\|_{\mathscr{F}} + 2E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i Pf(X_i)\Bigg\|_{\mathscr{F}} \\
&\quad + n^{-1/2} E\Bigg\| n^{-1} \sum_{j=1}^n Pf(X_j)\Bigg\| + E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i \pi_{2,2}^P f(X_i, X_j)\Bigg\|_{\mathscr{F}} \\
&\le O(n^{-1/2}) + 2E\Bigg\| n^{-1} \sum_{i=1}^n \varepsilon_i Pf(X_i)\Bigg\|_{\mathscr{F}} + E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i \pi_{2,2}^P f(X_i, X_j)\Bigg\|_{\mathscr{F}}.
\end{aligned}$$

By symmetrization, convexity and decoupling,

$$\begin{aligned}
E\Bigg\| n^{-1} \sum_{i=1}^n \varepsilon_i Pf(X_i)\Bigg\|_{\mathscr{F}} &\le O(n^{-1/2}) + E\Bigg\| n^{-1} \sum_{i=1}^n \varepsilon_i\big(Pf(X_i) - P^2 f\big)\Bigg\|_{\mathscr{F}} \\
&\le O(n^{-1/2}) + 2E\Bigg\| n^{-1} \sum_{i=1}^n \big(Pf(X_i) - P^2 f\big)\Bigg\|_{\mathscr{F}} \\
&\le O(n^{-1/2}) + 3E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \big(f(X_i, X_j') - P^2 f\big)\Bigg\|_{\mathscr{F}} \\
&\le O(n^{-1/2}) + cE\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \big(f(X_i, X_j) - P^2 f\big)\Bigg\|_{\mathscr{F}}.
\end{aligned}$$

We also have, again by decoupling and Jensen's inequality,

$$\begin{aligned}
E\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i \pi_{2,2}^P f(X_i, X_j)\Bigg\|_{\mathscr{F}} &\le cE\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \varepsilon_i \pi_{2,2}^P f(X_i, X_j')\Bigg\|_{\mathscr{F}} \\
&= cE\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \pi_{2,2}^P (f - P^2 f)(X_i, X_j')\Bigg\|_{\mathscr{F}} \\
&\le cE\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \big(f(X_i, X_j') - P^2 f\big)\Bigg\|_{\mathscr{F}} \\
&\le cE\Bigg\| n^{-2} \sum_{(i,j)\in I_2^n} \big(f(X_i, X_j) - P^2 f\big)\Bigg\|_{\mathscr{F}}.
\end{aligned}$$

(ii) is proved.

(ii) $\Rightarrow$ (iii). This part follows from a version of Sudakov's inequality for Rademacher processes [Carl and Pajor (1988); see Ledoux and Talagrand (1991), Corollary 4.14]: If $T \subset \mathbf{R}^N$ and $r(t) = E\sup_{t\in T}|\sum_{i=1}^N \varepsilon_i t_i|$, $t = (t_1,\dots,t_n)$, then, for all $\varepsilon > 0$,

$$\varepsilon\big(\log N(T, d_2, \varepsilon)\big)^{1/2} \le Kr(T)\left(\log\left(2 + \frac{N^{1/2}}{r(T)}\right)\right)^{1/2},$$

where d_2 is Euclidean distance. We apply this result to $n^{1/2}((n-m)!/n!)\cdot \sum_{I_m^n}\varepsilon_{i_1} f(X_{i_1},\dots,X_{i_m})$, $f \in \mathscr{F}$, noting that the d_2 distance for this process is precisely the $\tilde{e}_{n,2}$ distance. Hence,

$$\begin{aligned}\varepsilon\big(\log \tilde{N}_{n,2}(\varepsilon,\mathscr{F})\big)^{1/2} &\le n^{1/2}K\left(E_\varepsilon\left\|\frac{(n-m)!}{n!}\sum_{I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}}\right)\\ &\quad\times\left(\log\left(2 + \frac{1}{E_\varepsilon\|((n-m)!/n!)\sum_{I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\|_{\mathscr{F}}}\right)\right)^{1/2}.\end{aligned}$$

Since by (ii), $E_\varepsilon\|((n-m)!/n!)\sum_{I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\|_{\mathscr{F}} \to 0$ in probability, and $\tilde{e}_{n,1} \le \tilde{e}_{n,2}$, (iii) is proved.

(iii) $\Rightarrow$ (ii). Let $\mathscr{F}_M = \{fI_{F\le M}: f \in \mathscr{F}\}$. Since

$$\begin{aligned}&E\left\|\frac{(n-m)!}{n!}\sum_{I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}}\\ &\quad\le E\left\|\frac{(n-m)!}{n!}\sum_{I_m^n}\varepsilon_{i_1}(fI_{F\le M})(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}} + P^mFI_{F\ge M}\end{aligned}$$

and $P^mFI_{F\ge M} \to 0$ as $M \to \infty$, it suffices to show

$$\lim_{M\to\infty}\limsup_{n\to\infty} E\left\|\frac{(n-m)!}{n!}\sum_{I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}_M} = 0.$$

For ω fixed, given $\delta > 0$, let $\mathscr{F}^*$ be a subset of $\mathscr{F}_M$ such that $\#\mathscr{F}^* = \tilde{N}_{n,1}(\delta,\mathscr{F}_M)$ and $\sup_{f\in\mathscr{F}_M}\min_{f^*\in\mathscr{F}^*}\tilde{e}_{n,1}(f,f^*) \le \delta$. Then, by (2.14),

$$\begin{aligned}&\frac{(n-m)!}{n!}E_\varepsilon\left\|\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}_M}\\ &\quad\le \delta + KM\big(\log\tilde{N}_{n,1}(\delta,\mathscr{F}_M)\big)^{1/2}n^{-m}\\ &\qquad\times\max_{f\in\mathscr{F}^*}\left[\sum_{i_1}\left(\sum_{(i_2,\dots,i_m):\,(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m})\right)^2\right]^{1/2}\\ &\quad\le \delta + K\big(n^{-1}\log\tilde{N}_{n,1}(\delta,\mathscr{F}_M)\big)^{1/2}.\end{aligned}$$

Since

$$\tilde{e}_{n,1}(fI_{F\le M}, gI_{F\le M}) \le \tilde{e}_{n,1}(f,g) + 2U_m^n(FI_{F\ge M}),$$

it follows that if $U_m^n(FI_{F\ge M}) \le \delta/4$, then $\tilde{N}_{n,1}(\delta, \mathscr{F}_M) \le \tilde{N}_{n,1}(\delta/2, \mathscr{F})$. Hence

$$\Pr{}^*\{n^{-1}\log \tilde{N}_{n,1}(\delta, \mathscr{F}_M) \ge \varepsilon\}$$
$$\le \Pr{}^*\{n^{-1}\log \tilde{N}_{n,1}(\delta/2, \mathscr{F}) \ge \varepsilon\} + \Pr\{U_m^n(FI_{F>M}) > \delta/4\} \to 0$$

if $EFI_{F>M} < \delta/4$. Thus (ii) holds. □

We should remark that the preceding proof works with $\tilde{N}_{n,1}$ replaced by $\tilde{N}_{n,p}$ for any $p \in [1,2]$, and also by $\tilde{N}_{n,p}^{\mathrm{dec}}$, $p \in [1,2]$. In the case $F \le c < \infty$, an argument in Talagrand [(1987) page 863, reproduced in Dudley, Giné and Zinn (1991), page 503] shows that $n^{-1}\log \tilde{N}_{n,1}(\varepsilon, \mathscr{F}) \to 0$ in probability for all $\varepsilon > 0$ implies $n^{-1}\log \tilde{N}_{n,\infty}(\varepsilon, \mathscr{F}) \to 0$ in probability for all $\varepsilon > 0$; this is the only nontrivial part of the following statement: For $\mathscr{F}$ uniformly bounded, the conditions

$$\lim_{n\to\infty} n^{-1}\log \tilde{N}_{n,p}(\varepsilon, \mathscr{F}) = 0 \quad \text{in probability} \tag{3.2}$$

are all equivalent for $p \in (0,\infty]$, and likewise for $\tilde{N}_{n,p}^{\mathrm{dec}}$. So, in this case, (3.2) [or (3.2) decoupled] for any $0 < p \le \infty$ is also necessary and sufficient for the law of large numbers.

Since $\tilde{e}_{n,1}(S_m f, S_m g) \le e_{n,1}(f,g)$, we have the following corollary.

Corollary 3.2. *If $\mathscr{F}$ is a measurable class, then the conditions $P^m F < \infty$ and $\log N_{n,1}(\varepsilon, \mathscr{F})/n \to 0$ in probability* imply $\|U_m^n - P^m\|_{\mathscr{F}} \to 0$ a.s.*

The condition $n^{-1}\log N_{n,1}(\varepsilon, \mathscr{F}) \to_{\Pr} 0$ for all $\varepsilon > 0$ is stronger than the condition $n^{-1}\log \tilde{N}_{n,1}(\varepsilon, \mathscr{F}) \to_{\Pr} 0$ for all $\varepsilon > 0$. To show this, we exhibit a class of functions that satisfies the first condition and not the second (for simplicity, in the case $m = 2$). First we note that the part (iii) ⇒ (ii) in the proof of Theorem 3.1 shows that $n^{-1}\log N_{n,1}(\varepsilon, \mathscr{F}) \to_{\Pr} 0$ for all $\varepsilon > 0$ implies that $E\|n^{-2}\sum_{I_2^n}\varepsilon_i'|f(X_i,X_j)|\,\|_{\mathscr{F}} \to 0$. Hence, it is enough to find a class $\mathscr{F}$ such that $E\|n^{-2}\sum_{I_2^n} f(X_i,X_j)\|_{\mathscr{F}} \to 0$, but $E\|n^{-2}\sum_{I_2^n}\varepsilon_i'|f(X_i,X_j)|\,\|_{\mathscr{F}} \not\to 0$. We take $S = \{-1,1\}\times[0,1]$ with the product measure of $(1/2)\delta_{-1} + (1/2)\delta_1$ and Lebesgue measure, and $X = (\varepsilon, Y)$. Let $\mathscr{H} = \{h\colon [0,1] \to [-1,1]\colon h$ is Borel measurable$\}$ and $\mathscr{F} = \{f\colon S \to \mathbb{R}\colon f(x_1,x_2) = \varepsilon_1\varepsilon_2(h(y_1)+h(y_2))$ for some $h \in \mathscr{H}\}$. Then an argument in the proof of Theorem 3.1 [the last part of (i) ⇒ (ii)] gives $E\|n^{-2}\sum_{I_2^n}\varepsilon_i' f(X_i,X_j)\|_{\mathscr{F}} \le cE\|n^{-2}\sum_{I_2^n} f(X_i,X_j)\|_{\mathscr{F}}$; hence, since

$$E\Big\|n^{-2}\sum_{I_2^n} f(X_i,X_j)\Big\|_{\mathscr{F}} \le 2E\Big|n^{-1}\sum_{i=1}^n \varepsilon_i\Big|\,\Big\|n^{-1}\sum_{j=1}^n \varepsilon_j h(Y_j)\Big\|_{\mathscr{H}}$$
$$+\, E\Big\|2n^{-2}\sum_{i=1}^n h(Y_i)\Big\|_{\mathscr{H}} \to 0,$$

it follows that $E\|n^{-2}\sum_{I_2^n}\varepsilon_i' f(X_i, X_j)\|_{\mathscr{F}} \to 0$. We also have that

$$E\left\|n^{-2}\sum_{I_2^n}\varepsilon_i'|f(X_i,X_j)|\right\|_{\mathscr{F}} \geq E\left\|n^{-2}\sum_{i,j=1}^{n}\varepsilon_i'|h(Y_i)+h(Y_j)|\right\|_{\mathscr{H}} - E\left\|n^{-2}\sum_{i=1}^{n}\varepsilon_i'|f(X_i,X_i)|\right\|_{\mathscr{F}}.$$

Conditionally on ε' and Y, we take a function $h \in \mathscr{H}$ such that $h(Y_i) = (1 + \varepsilon_i')/2$. Then $n^{-2}\sum_{i,j=1}^{n}\varepsilon_i'|h(Y_i) + h(Y_j)| = 2^{-1}(1 + n^{-1}\sum_{i=1}^{n}\varepsilon_i')^2$. Therefore, $E\|n^{-2}\sum_{I_2^n}\varepsilon_i'|f(X_i, X_j)|\|_{\mathscr{F}} \nrightarrow 0$. So, $\mathscr{F}$ has the desired properties.

Corollary 3.2 directly gives, by Pollard [(1984), page 27], that:

Corollary 3.3. *If $\mathscr{F}$ is a measurable VC-subgraph class of functions with $P^mF < \infty$, then $\|U_m^n - P^m\|_{\mathscr{F}} \to 0$ a.s.*

For $m = 2$ this was already observed in Nolan and Pollard (1987).

Example 3.4. Corollary 3.3 provides another proof of Theorem 2.2 in Helmers, Janssen and Serfling (1988): If h: $S^m \to \mathbb{R}$ is a Borel symmetric function and if q: $(0,1) \to \mathbb{R}^+$ is continuous, nondecreasing on $(0,\delta]$ and nonincreasing on $[1-\delta, 1)$, for some $0 < \delta < 1/2$, and satisfies $\int_0^1 dt/q(t) < \infty$, then

$$\frac{(n-m)!}{n!}\left\|\sum_{I_m^n}\frac{I_{h(X_{i_1},\ldots,X_{i_m})\leq t} - P^m(h\leq t)}{q(P^m(h\leq t))}\right\|_{\infty} \to 0 \quad \text{a.s.}$$

In fact the class $\{(I_{h\leq t} - P(h \leq t))/q(P(h \leq t)) : t \in \mathbb{R}\}$ is a VC-subgraph class and the integrability condition on q gives $P^mF < \infty$.

As another example we obtain the analogue of the Blum–DeHardt law of large numbers [e.g., Dudley (1984)] for U-statistics. A direct proof is also very easy to obtain. We recall that $N_{[]}^{(p)}(\varepsilon, \mathscr{F}, P^m) = \min\{r$: there exist $f_1, \ldots, f_r$ and $\Delta_1, \ldots, \Delta_r \in L_p(P^m)$ s.t. $P^m|\Delta_i|^p < \varepsilon^p$ and for all $f \in \mathscr{F}$ there exists $i \leq r$ with $|f_i - f| < \Delta_i\}$. The set of f's such that $|f_i - f| < \Delta_i$ is called the ith bracket A_i.

Corollary 3.5. *Let $\mathscr{F}$ be a measurable class of functions on S^m such that, for all $\varepsilon > 0$, $N_{[]}^{(1)}(\varepsilon, \mathscr{F}, P^m) < \infty$. Then $\|U_m^n - P^m\|_{\mathscr{F}} \to 0$ a.s.*

Proof. For $\varepsilon > 0$ let $\{f_j\}$ and $\{\Delta_j\}$ be a set of functions as in the definition of $N_{[]}^{(1)}(\varepsilon, \mathscr{F}, P^m)$. Then if $P^m\Delta_j < \varepsilon$ we have

$$((n-m)!/n!)\sum_{(i_1,\ldots,i_m)\in I_m^n}\Delta_j(X_{i_1},\ldots,X_{i_m}) \leq \varepsilon$$

for all $n > n_0(\omega)$, for some $n_0(\omega) < \infty$ a.s. by the law of the large numbers for

U-statistics. Hence, for each $n > n_0(\omega)$, $N_{n,1}(\varepsilon, \mathscr{F}) \le N_{[]}^{(1)}(\varepsilon, \mathscr{F}, P^m)$. Now Corollary 3.2 gives the result. □

Dudley [(1984), Proposition 6.1.7] shows that if $\mathscr{F}$ is the unit ball of the dual of a separable Banach space B and Q is a Borel probability measure on B with $\int \|x\|\, dQ < \infty$, then $N_{[]}^{(1)}(\varepsilon, \mathscr{F}, Q) < \infty$ for all $\varepsilon > 0$. Let $H\colon S^m \to B$ satisfy $P^m\|H\| < \infty$, and let $Q = P^m \circ H^{-1}$. Then Dudley's observation, together with Corollary 3.5 applied to $\mathscr{F} = \{f \circ H\colon f \in B_1'\}$, immediately gives the following theorem.

THEOREM 3.6 (Law of large numbers for B-valued U-statistics). *Let B be a separable Banach space and let $H\colon S^m \to B$ be a measurable function such that $P^m\|H\| < \infty$. Then*

$$\|U_m^n(H) - P^m(H)\| \to 0 \quad a.s.$$

EXAMPLE 3.7. Helmers, Janssen and Serfling (1988) consider a.s. convergence in $L_p(\mathbb{R}, \mathscr{B}, \lambda)$ $p \ge 1$, of the random process $U_m^n(I_{h \le t}, P) - P^m(h \le t)$, $t \in \mathbb{R}$, for $h\colon S^m \to \mathbb{R}$ symmetric. We prove the following: For $p \ge 1$,

$$E|h|^{1/p} < \infty \quad \text{implies } \|U_m^n(I_{h \le t}) - P^m\{h \le t\}\|_p \to 0 \quad \text{a.s.}$$

As a consequence of Theorem 3.6 it suffices to show that

$$E\left(\int |I_{h \le t} - P^m\{h \le t\}|^p \, d\lambda\right)^{1/p} < \infty.$$

By Jensen's inequality in $L_p[0, \infty)$,

$$\left(\int_0^\infty (P^m\{|h| > t\})^p \, dt\right)^{1/p} = \|EI_{|h|>t}\|_p \le E\|I_{|h|>t}\|_p = E|h|^{1/p} < \infty.$$

Then $\int_{-\infty}^\infty |I_{h \le t} - P^m\{h \le t\}|^p \, dt \le 2\int_0^\infty (P^m\{|h| > t\})^p \, dt + |h|$ and the result follows.

Conversely, for $m = 1$, by the LLN, if $\|U_1^n(I_{h \le t}) - P\{h \le t\}\|_p \to 0$ a.s., then $E(\int |I_{h \le t} - P\{h \le t\}|^p \, d\lambda)^{1/p} < \infty$. Since $\|U_1^n(I_{h \le t}) - P\{h \le t\}\|_p < \infty$ implies $\int_0^\infty (P\{|h| > t\})^p \, dt < \infty$ and since, for $h \ge 0$,

$$\int_{-\infty}^\infty |I_{h \le t} - P\{h \le t\}|^p \, dt = \int_0^\infty (P\{h > t\})^p + (P\{h \le -t\})^p \, dt$$
$$+ \int_0^h \left((P\{h \le t\})^p - (P\{h \ge t\})^p\right) dt$$

(and a similar identity holds for $h \le 0$), we have

$$\lim_{|h| \to \infty} \int_0^\infty |I_{h \le t} - P\{h \le t\}|^p \, dt / |h| = 1 \quad \text{and} \quad E|h|^{1/p} < \infty.$$

EXAMPLE 3.8. If P is discrete, then any class of functions $\mathscr{F}$ on S^m such that $P^m F < \infty$ satisfies the law of large numbers.

PROOF. We can assume $S = \mathbf{N}$. For $r \in \mathbb{N}$, let $T_r = \{(n_1, \dots, n_m) \in \mathbb{N}^m: n_i \le r\}$. Given $\varepsilon > 0$, let r be such that $\int_{T_r^c} F\,dP < \varepsilon/4$ and let $\mathscr{G} = \{f: S^m \to \mathbb{R},\ f|_{T_r^c} \equiv F|_{T_r^c}$ or $\equiv -F|_{T_r^c}$ and, for $x \in T_r$, $f(x) = k\varepsilon/2$ for some integer k such that $-1 - 2F(x)/\varepsilon \le k \le 1 + 2F(x)/\varepsilon\}$. Then $\#\mathscr{G} < \infty$ and for any $f \in \mathscr{F}$ there are two functions in $\mathscr{G}$, say g_1 and g_2, such that $g_1 \le f \le g_2$, $g_2 - g_1|_{T_r} \le \varepsilon/2$ and $g_2 - g_1|_{T_r^c} \equiv 2F|_{T_r^c}$. Hence, Corollary 3.5 applies. □

EXAMPLE 3.9. If P is Lebesgue measure (or has a density f with respect to Lebesgue measure which is bounded and bounded away from zero), then the class of all the indicator functions of convex sets in $[0,1]^m$ satisfies the law of large numbers.

PROOF. Bronštein's theorem (1976) [see also Dudley (1984), 7.3.2] states that there is a constant c such that $\log N_{[\,]}^{(1)}(\varepsilon, \mathscr{F}, P^m) \le c\varepsilon^{(1-m)/2}$. Hence Example 3.9 follows from Corollary 3.5. □

EXAMPLE 3.10. Let $F: \mathbb{R}^m \to \mathbb{R}$, $F > 0$, $P^m F < \infty$. Let, for $0 < \alpha \le 1$ and $c < \infty$, $\mathscr{F} = \{g: \mathbb{R}^m \to \mathbb{R}: |g(x)| \le F(x)$ for all $x \in \mathbb{R}^m$, $|g(x) - g(y)| \le c|x - y|^\alpha$ for all $x, y \in \mathbb{R}^m\}$. Then $\|U_m^n - P^m\|_{\mathscr{F}} \to 0$ a.s.

PROOF. Let $\Phi_M(x) = x \wedge M \vee (-M)$, $M > 0$. Then

$$|f - \Phi_M \circ f| \le F I_{F > M} = F^M$$

so that $\|(U_m^n - P^m)(f - \Phi_M \circ f)\|_{\mathscr{F}} \le (U_m^n + P^m)F^M \to 0$ (as $n \to \infty$ and then $M \to \infty$). So, since the modulus of continuity of $\Phi_M \circ f$ is not larger than that of f, we may assume $\mathscr{F}$ is uniformly bounded by 1. Let R_l be the coordinate hypercube of side l centered at 0. Then $\|(U_m^n - P^m) f I_{R_l^c}\|_{\mathscr{F}} \le c(U_m^n + P^m) F I_{R_l^c} \to 0$ (as $n \to \infty$ and then $l \to \infty$). So we may further assume that the functions in $\mathscr{F}$ are supported by a fixed compact set. But then, by a result of Kolmogorov [Theorem 7.1.1 in Dudley (1984)] the metric entropy of $\mathscr{F}$ with respect to the sup norm is of the order of $\varepsilon^{-m/\alpha}$, hence $N_{[\,]}^{(1)}(\varepsilon, \mathscr{F})$ is at most of the same order and the law of the large numbers follows from Corollary 3.5. □

Finally, we consider the LLN for V-processes $\{V_m^n(f, P) = P_n^m f: f \in \mathscr{F}\}$. By decomposition into U-processes and Marcinkiewicz type laws of large numbers for the diagonals [Sen (1974); see also Giné and Zinn (1992a)] we obtain the following theorem. Here and elsewhere, given a set C, $\#C$ will denote the cardinality of C.

THEOREM 3.11. *If $\mathscr{F}$ is measurable with envelope F satisfying the integrability conditions*

$$E\left|F(X_{i_1}, \dots, X_{i_m})\right|^{\#\{i_1, \dots, i_m\}/m} < \infty,$$

$\{V_m^n(f): f \in \mathscr{F}\}$ verifies the LLN if and only if $\{U_m^n(f): f \in \mathscr{F}\}$ does.

PROOF. We only consider $m = 2$; the general case is similar. We have

$$\left\| n^{-2} \sum_{i,j=1}^{n} \left(f(X_i, X_j) - P^2 f \right) - \frac{n-1}{n} \left(U_2^n(f) - P^2 f \right) \right\|_{\mathscr{F}}$$

$$\le n^{-2} \sum_{i=1}^{n} F(X_i, X_i).$$

But by the Marcinkiewicz laws of large numbers, $n^{-2}\sum_{i=1}^{n} F(X_i, X_i) \to 0$ a.s. $\square$

4. The central limit theorem for nondegenerate U-processes. As in the case of a single function f, to prove the CLT for $\{U_m^n(f): f \in \mathscr{F}\}$ with a Gaussian limit, we must prove that $\{n^{1/2} U_1^n(\pi_{1,m}^P f): f \in \mathscr{F}\}$ converges in $l^\infty(\mathscr{F})$ to a Gaussian process and that the processes $\|n^{1/2} U_k^n(\pi_{k,m}^P f)\|_{\mathscr{F}} \to 0$ in probability for $1 < k \le m$. The first condition is equivalent to the class $\{\pi_{1,m}^P f: f \in \mathscr{F}\}$ being P-Donsker, a question that has been thoroughly studied. Therefore only the second condition must be dealt with. Note that the CLT holds for $\mathscr{F}$ if and only if it holds for $S_m \mathscr{F}$. So, in this section and in Section 5, only classes of symmetric functions are considered. In what follows, given a probability measure P on $(S, \mathscr{S})$, d denotes the pseudodistance on $\mathscr{F}$ given by

$$d^2(f, g) = P\left(P^{m-1}(f - g)\right)^2,$$

and then we define

$$\mathscr{F}_\delta' = \{ f - g: f, g \in \mathscr{F}, d(f, g) \le \delta \}$$

for all $\delta > 0$.

We begin with a general result similar to Theorem 2.8 in Giné and Zinn (1986).

THEOREM 4.1. *Let $\mathscr{F}$ be a measurable class of symmetric functions on S^m such that $t^2 \Pr\{F > t\} \to 0$. Then the following are equivalent*:

(a) *$\mathscr{F}$ satisfies the CLT*:

$$\{ n^{1/2}(U_m^n(f, P) - P^m) : f \in \mathscr{F} \} \to_{\mathscr{L}} \{ m G_P \circ P^{m-1} f : f \in \mathscr{F} \}. \tag{4.1}$$

(b) *$(\mathscr{F}, d)$ is totally bounded and*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} E \left\| n^{1/2}(U_m^n(f, P) - P^m f) \right\|_{\mathscr{F}_\delta'}^r = 0 \tag{4.2}$$

for some (*all*) $0 < r < 2$.

(c) $(\mathscr{F}, d)$ *is totally bounded and*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr\left\{\left\|n^{-m+1/2}\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}'_\delta}\ge\lambda\right\}=0 \tag{4.3}$$

for all $\lambda > 0$.

(d) $(\mathscr{F}, d)$ *is totally bounded and*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}E\left\|n^{-m+1/2}\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}'_\delta}^r=0 \tag{4.4}$$

for some (*all*) $0 < r < 2$.

PROOF. As observed in the introduction, (a) is equivalent to:

(b′) $(\mathscr{F}, d)$ is totally bounded and

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr\left\{\left\|n^{1/2}(U_m^n(f,P)-P^mf)\right\|_{\mathscr{F}'_\delta}\ge\lambda\right\}=0$$

for all $\lambda > 0$.

The equivalence between (d) and (b) is contained in the proof of Theorem 3.1. Hence, we need only prove (c) implies (d) and (b′) implies (b). Both proofs are similar, so we only prove the former. It suffices to show that the sequence

$$\left\{\left\|n^{-m+1/2}\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}'_\delta}^r\right\}$$

is uniformly integrable for small δ. This will follow if we show

$$\sup_n E\left\|n^{-m+1/2}\sum_{(i_1,\dots,i_m)\in I_m^n}\varepsilon_{i_1}f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}'_\delta}^p<\infty \tag{4.5}$$

for some $p > r$, $1 < p < 2$. For simplicity we prove (4.5) only for $m = 2$, the general case being similar. The constant c in the following inequalities may vary from line to line. We have, as in the proof of (i) $\Rightarrow$ (ii) in Theorem 3.1,

$$\begin{aligned}
&E\left\|n^{-3/2}\sum_{(i,j)\in I_2^n}\varepsilon_i f(X_i,X_j)\right\|_{\mathscr{F}}^p\\
&\quad\le E\left\|n^{-3/2}\sum_{(i,j)\in I_2^n}\varepsilon_i\pi_{2,2}^P f(X_i,X_j)\right\|_{\mathscr{F}}^p\\
&\qquad+2E\left\|n^{-1/2}\sum_{i=1}^n\varepsilon_i Pf(X_i)\right\|_{\mathscr{F}}^p+O(1).
\end{aligned}$$

Now,

$$E\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i \pi_{2,2}^P f(X_i, X_j)\right\|_{\mathscr{F}}^p$$

$$\le cE\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i \pi_{2,2}^P f(X_i, X_j')\right\|_{\mathscr{F}}^p \qquad \text{by Proposition 2.1}$$

$$\le cE\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i \varepsilon_j' \pi_{2,2}^P f(X_i, X_j')\right\|_{\mathscr{F}}^p \qquad \text{by the usual randomization}$$

$$\le c\left(E\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i \varepsilon_j' \pi_{2,2}^P f(X_i, X_j')\right\|_{\mathscr{F}}\right)^p$$

$$+ cE\max_{i\le n}\left\| n^{-3/2} \sum_{j:\, j\ne i,\, j\le n} \varepsilon_j' \pi_{2,2}^P f(X_i, X_j')\right\|_{\mathscr{F}}^p \qquad \text{by Proposition 2.8}$$

$$\le c\left(E\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i \pi_{2,2}^P f(X_i, X_j')\right\|_{\mathscr{F}}\right)^p + O(1)$$

by randomization and the tail condition on F

$$\left(E\max_{i\le n}\|\cdots\|_{\mathscr{F}}^p \le cE\left(n^{-3/2}\sum_{j=1}^n \max_{i\le n} F(X_i, X_j')\right)^p \right.$$

$$\left. \le cE\left(\max_{i\le n} F(X_i, X_1')/n^{1/2}\right)^p \to 0\right)$$

$$\le c\left(E\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i f(X_i, X_j')\right\|_{\mathscr{F}}\right)^p + O(1)$$

by Jensen's inequality.

By Hoffmann-Jørgensen's inequality (Proposition 2.8 with $m = 1$), the tail condition on F and Jensen's inequality,

$$E\left\| n^{-1/2}\sum_{i=1}^n \varepsilon_i Pf(X_i)\right\|_{\mathscr{F}}^p \le c\left(E\left\| n^{-1/2}\sum_{i=1}^n \varepsilon_i Pf(X_i)\right\|_{\mathscr{F}}\right)^p + O(1)$$

$$\le c\left(E\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i f(X_i, X_j')\right\|_{\mathscr{F}}\right)^p + O(1).$$

Therefore, by an argument in the first part of the proof of Theorem 3.1, we have

$$E\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i f(X_i, X_j)\right\|_{\mathscr{F}}^p \le \left(E\left\| n^{-3/2} \sum_{(i,j)\in I_2^n} \varepsilon_i f(X_i, X_j)\right\|_{\mathscr{F}}\right)^p + O(1).$$

For δ small, this, (4.3) and Paley–Zygmund's inequality [Kahane (1968), page 6] imply (4.5) and (d) follows. Since this proof involves only expected values (except for the last step which also involves probabilities), it is clear that the proof of (b′) ⇒ (b) is obtained by combining the present proof and that of (i) ⇔ (ii) in Theorem 3.1. □

Note that we also have that (c) and (d) are equivalent to the corresponding decoupled (c) and (d). Here is an interesting consequence:

COROLLARY 4.2. *Let $\mathscr{F} \subset L^2(S^m, \mathscr{S}^m, P^m)$ be a measurable class of symmetric functions such that $t^2 P^m\{F > t\} \to 0$. Then the following are equivalent:*

(a) *$\mathscr{F}$ satisfies the CLT in* (4.1).

(b) *$P^{m-1}\mathscr{F}$ is a P-Donsker class and $\|n^{1/2}U_k^n(\pi_{k,m} f)\|_{\mathscr{F}} \to 0$ in probability and (or) in L_r for all (some) $0 < r < 2$, $k = 2, \dots, m$.*

(c) *$P^{m-1}\mathscr{F}$ is a P-Donsker class and*

$$E\left\| n^{-m+1/2} \sum_{(i_1, \dots, i_m) \in I_m^n} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} f(X_{i_1}, \dots, X_{i_m}) \right\|_{\mathscr{F}}^r \to 0$$

for all (some) $0 < r < 2$.

In fact, (c) or (b) imply (a) without requiring any tail conditions on F.

PROOF. It is obvious that (b) implies (a). We prove first that (a) implies (b). The proof reduces to showing asymptotic equicontinuity for

$$\{n^{1/2}U_k^n(\pi_{k,m}^P f)\}_{n=1}^{\infty}, \qquad k = 1, \dots, m,$$

assuming (a). By Jensen's inequality and decoupling, for $k = 1, \dots, m$,

$$E\left\| n^{-k+1/2} \sum_{I_k^n} \varepsilon_{i_1} \pi_{k,m}^P f(X_{i_1}, \dots, X_{i_k}) \right\|_{\mathscr{F}'_\delta}$$
$$\leq 2^k E\left\| n^{-m+1/2} \sum_{I_m^n} \varepsilon_{i_1} f(X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)}) \right\|_{\mathscr{F}'_\delta}.$$

Now Theorem 4.1 (decoupled version) gives the desired asymptotic equicontinuity. Hence (b) holds.

Now, assume (b). We prove (c) in the case $m = 2$ (the general case is similar). We have

$$E\left\| n^{-3/2} \sum_{(i,j) \in I_2^n} \varepsilon_i \varepsilon'_j f(X_i, X_j) \right\|_{\mathscr{F}} \leq E\left\| n^{-3/2} \sum_{(i,j) \in I_2^n} \varepsilon_i \varepsilon'_j \pi_{2,2}^P f(X_i, X_j) \right\|_{\mathscr{F}}$$
$$+ 2E\left\| n^{-3/2} \sum_{(i,j) \in I_2^n} \varepsilon_i \varepsilon'_j \pi_{1,2}^P f(X_i) \right\|_{\mathscr{F}}$$
$$+ E\left\| n^{-3/2} \sum_{(i,j) \in I_2} \varepsilon_i \varepsilon'_j P^2 f \right\|_{\mathscr{F}}$$

which goes to 0 by (b), symmetrization and decoupling as in previous proofs:

An argument similar to the one in the last part of the proof of Theorem 4.1 (i.e., Proposition 2.8 and Paley–Zygmund) shows that if $\|n^{1/2}U_k^n(\pi_{k,m}f)\|_{\mathscr{F}} \to 0$ in probability, then $E\|n^{1/2}U_k^n(\pi_{k,m}f)\|_{\mathscr{F}}^r \to 0$ for all $0 < r < 2$. Now we can decouple the first summand in the preceding inequality, cancel the ε_i's and ε_j''s by the usual symmetrization for sums for independent random variables, and then "undecouple" to get it dominated by $\|n^{1/2}U_2^n(\pi_{2,2}f)\|_{\mathscr{F}}$, which tends to 0 by (b). For the second summand, we insert the diagonal, factorize $n^{-1/2}\sum_{j=1}^n \varepsilon_j'$ and compare with $n^{-1}U_1^n(\pi_{1,2}f)$. The third summand tends to 0 because $\|P^2 f\|_{\mathscr{F}} < \infty$ and $n^{-3/2}\sum_{I_2^n}\varepsilon_i\varepsilon_j' \to 0$.

Finally we will show that (c) implies (b). We have, for $2 \le k \le m$,

$$
\begin{aligned}
&E\Bigg\| n^{-k+1/2} \sum_{(i_1,\dots,i_k)\in I_k^n} \pi_{k,m}^P f(X_{i_1},\dots,X_{i_k})\Bigg\|_{\mathscr{F}} \\
&\le cE\Bigg\| n^{-k+1/2} \sum_{(i_1,\dots,i_k)\in I_k^n} \pi_{k,m}^P f\big(X_{i_1}^{(1)},\dots,X_{i_k}^{(k)}\big)\Bigg\|_{\mathscr{F}} \qquad \text{by decoupling} \\
&\le cE\Bigg\| n^{-k+1/2} \sum_{(i_1,\dots,i_k)\in I_k^n} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)}\pi_{k,m}^P f\big(X_{i_1}^{(1)},\dots,X_{i_k}^{(k)}\big)\Bigg\| \\
&\qquad \text{by symmetrization} \\
&\le cE\Bigg\| n^{-m+1/2} \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} f\big(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\big)\Bigg\|_{\mathscr{F}} \\
&\qquad \text{by Jensen's inequality, repeatedly} \\
&\le cE\Bigg\| n^{-m+1/2} \sum_{(i_1,\dots,i_m)\in I_m^n} \sum_{1\le j_1<j_2\le m} \varepsilon_{i_{j_1}}^{(j_1)}\varepsilon_{i_{j_2}}^{(j_2)} f\big(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)}\big)\Bigg\|_{\mathscr{F}} \\
&\qquad \text{by Jensen's inequality} \\
&\le cE\Bigg\| n^{-m+1/2} \sum_{(i_1,\dots,i_m)\in I_m^n} \sum_{1\le j_1<j_2\le m} \varepsilon_{i_{j_1}}\varepsilon_{i_{j_2}} f(X_{i_1},\dots,X_{i_m})\Bigg\|_{\mathscr{F}} \\
&\qquad \text{by decoupling} \left(\text{note that the function } \sum_{1\le j<k\le m} y_j y_k f(x_1,\dots,x_m) \text{ of } (x_1,y_1),\dots,(x_m,y_m) \text{ is symmetric}\right) \\
&\le cE\Bigg\| n^{-m+1/2} \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1}\varepsilon_{i_2} f(X_{i_1},\dots,X_{i_m})\Bigg\|_{\mathscr{F}} \\
&\le cE\Bigg\| n^{-m+1/2} \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} f(X_{i_1},\dots,X_{i_m})\Bigg\|_{\mathscr{F}} \qquad \text{by decoupling.}
\end{aligned}
$$

The result follows. $\square$

(a) *The case of uniformly bounded $\mathscr{F}$.* The following is an analogue for U-processes of Theorem 2.1.1 in Giné and Zinn (1986).

THEOREM 4.3. *Let $\mathscr{F}$ be a measurable class of uniformly bounded symmetric functions on S^m. If*
(i) *$P^{m-1}\mathscr{F}$ is P-pre-Gaussian and*
(ii) *for some $\varepsilon > 0$,*

$$\left\| n^{1/2-m} \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1} f(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{F}'_{\varepsilon n^{-1/4}}} \to 0 \quad \text{in pr}^*, \tag{4.6}$$

then $\mathscr{F}$ satisfies the CLT in (4.1).

PROOF. Only the changes from the proof in Giné and Zinn [(1986), pages 78–80] are given. Since $P^{m-1}\mathscr{F}$ is P-pre-Gaussian and $\|P^m f\|_{\mathscr{F}} < \infty$, $(\mathscr{F}, d)$ is totally bounded, where $d^2(f,g) = P(P^{m-1}(f-g))^2$. Let $\mathscr{H}$ be a maximal set of functions h_i d-separated by more than $\varepsilon n^{-1/4}$. Only the computation of

$$\Pr\left\{ \sup_{f\in\mathscr{H}'-\{0\}} n^{1-2m} \sum_{i_1=1}^{n} \left(\sum_{(i_2,\dots,i_m):\,(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m}) \right)^2 \geq 4Ef(X_1,X_2,\dots,X_m)\,f(X_1,X_2',\dots,X_m') \right\}$$

requires a comment, since the rest follows in complete analogy with the preceding reference. We first note

$$\begin{aligned}
&\Pr\left\{ \sup_{f\in\mathscr{H}'-\{0\}} n^{1-2m} \sum_{i_1=1}^{n} \left(\sum_{(i_2,\dots,i_m):\,(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m}) \right)^2 \geq 4Ef(X_1,X_2,\dots,X_m)\,f(X_1,X_2',\dots,X_m') \right\} \\
&\quad \leq (\#\mathscr{H})^2 \sup_{f\in\mathscr{H}'-\{0\}} \Pr\left\{ n^{1-2m} \sum_{i_1=1}^{n} \left(\sum_{(i_2,\dots,i_m):\,(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m}) \right)^2 \geq 4Ef(X_1,X_2,\dots,X_m)\,f(X_1,X_2',\dots,X_m') \right\}.
\end{aligned}$$

Since the class $\mathscr{F}$ is uniformly bounded,

$$\begin{aligned}
&n^{1-2m} \sum_{i_1=1}^{n} \left(\sum_{(i_2,\dots,i_m):\,(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m}) \right)^2 \\
&\qquad - n^{1-2m} \sum_{(i_1,\dots,i_{2m-1})\in I_{2m-1}^n} f(X_{i_1},\dots,X_{i_m})\,f(X_{i_1},X_{i_{m+1}},\dots,X_{i_{2m-1}}) \\
&\quad = O(n^{-1}) < \varepsilon n^{-1/2} < P(P^{m-1}f)^2.
\end{aligned}$$

Hence, by the inequality in Proposition 2.3(a) and by Sudakov's inequality,

$$
\begin{aligned}
&(\#\mathscr{H})^2 \sup_{f\in\mathscr{H}'-\{0\}} \Pr\left\{ n^{1-2m} \sum_{i_1=1}^{n} \left(\sum_{(i_2,\dots,i_m):\,(i_1,\dots,i_m)\in I_m^n} f(X_{i_1},\dots,X_{i_m}) \right)^2 \right. \\
&\qquad\qquad \left. \geq 4Ef(X_1,X_2,\dots,X_m)\, f(X_1,X_2',\dots,X_m') \right\} \\
&\quad \leq (\#\mathscr{H})^2 \sup_{f\in\mathscr{H}'-\{0\}} \Pr\left\{ n^{1-2m} \sum_{I_{2m-1}^n} \big(f(X_{i_1},\dots,X_{i_m}) \right. \\
&\qquad\qquad \times f(X_{i_1},X_{i_{m+1}},\dots,X_{i_{2m-1}}) \\
&\qquad\qquad - Ef(X_1,X_2,\dots,X_m)\, f(X_1,X_2',\dots,X_m') \big) \\
&\qquad\qquad \left. \geq 2Ef(X_1,X_2,\dots,X_m)\, f(X_1,X_2',\dots,X_m') \right\} \\
&\quad \leq \exp\left(2c_n n^{1/2}\varepsilon^{-2} - 3\left[\frac{n}{2m-1}\right]\varepsilon^2 n^{-1/2}/2 \right).
\end{aligned}
$$

Now (4.6) follows as in the reference. □

Although Theorem 4.3 is theoretically interesting, Corollary 4.2 is more useful in what follows.

THEOREM 4.4. *Let $\mathscr{F}$ be a uniformly bounded class of real symmetric functions on S^m such that $P^{m-1}\mathscr{F}$ is a P-Donsker class and*

$$
\lim_{n\to\infty} E^*\left[n^{-1/2}\log N_{n,1}(\delta n^{-1/2},\mathscr{F})\right] = 0 \tag{4.7}
$$

for all $\delta > 0$. Then

$$
n^{1/2}(U_m^n - P^m) \to_{\mathscr{L}} mG_P\circ P^{m-1} \quad \text{in } l^\infty(\mathscr{F}).
$$

PROOF. Using inequality (2.10), letting $\mathscr{H}_n$ be a minimal $\delta n^{-1/2}$ dense set of $(\mathscr{F}, e_{n,1})$, we have

$$
\begin{aligned}
&E\left\| n^{-m+1/2} \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} f(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{F}} \\
&\quad \leq c\delta + cE^*\left\| n^{1/2-m} \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} f(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{H}_n} \\
&\quad \leq c\delta + cE^*\left[\log N_{n,1}(\delta n^{-1/2},\mathscr{F}) \right. \\
&\qquad\qquad \left. \times \max_{f\in\mathscr{H}_n} E_\varepsilon \left| n^{1/2-m} \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} f(X_{i_1},\dots,X_{i_m}) \right| \right] \\
&\quad \leq c\delta + cE^*\big(n^{-1/2}\log N_{n,1}(\delta n^{-1/2},\mathscr{F})\big).
\end{aligned}
$$

Hence Theorem 4.4 follows from Corollary 4.2. □

The following proposition generalizes Proposition 4.3.1 in Giné and Zinn (1986) [see Dudley (1984), for the same result for classes of sets].

PROPOSITION 4.5. *Let* $\mathscr{F} = \{f_k\}$. *Then* $\sup_k \|f_k\| < \infty$ *and* $\Sigma_{k=1}^{\infty}(E|f_k|^r)^s < \infty$ *for some* $r, s > 0$ *implies that* $\mathscr{F}$ *satisfies the CLT, that is,*

$$n^{1/2}(U_m^n - P^m) \to_{\mathscr{L}} mG_P \circ P^{m-1} \circ S_m \quad in\ l^\infty(\mathscr{F}). \tag{4.8}$$

PROOF. It is easy to see that we may assume $r = 2$ and the functions f_k symmetric. Also $\|P^{m-1}f_j\|_\infty \le \|f_j\|_\infty$ and $\|P^{m-1}f_j\|_2 \le \|f_j\|_2$. So the class $P^{m-1}\mathscr{F}$ satisfies the CLT (see the above reference).

The argument in Giné and Zinn (1986), using Proposition 2.4 instead of their bound for binomial probabilities, gives $\Pr\{n^{-1/2}\log N_{n,1}(\delta n^{-1/2}, \mathscr{F}) \ge \varepsilon\} \le e^{-cn}$ for some $c > 0$ and n large enough. By considering $(f(X_{i_1}, \dots, X_{i_m})$: $(i_1, \dots, i_m) \in I_m^n)$ as points of $[0,1]^{n!/(n-m)!}$ it is easy to see that $N_{n,1}(\varepsilon, \mathscr{F}) \le (1/\varepsilon)^{n^m}$. Hence $n^{-1/2}\log N_{n,1}(\delta n^{-1/2}, \mathscr{F}) \le n^{m-1/2}\log(n^{1/2}\delta^{-1}) \le n^m$ for n large. It follows that $E^*(n^{-1/2}\log N_{n,1}(\delta n^{-1/2}, \mathscr{F})) \le \varepsilon + n^m e^{-cn}$. Now Theorem 4.4 gives the result. □

THEOREM 4.6. *Let* $\mathscr{F}$ *be a measurable uniformly bounded class of real functions on* S^m *and let* P *be a probability measure on* $(S, \mathscr{S})$ *such that* $P^{m-1}\mathscr{F}$ *is measurable and P-pre-Gaussian. Then* $\varepsilon \log N_{[]}^{(1)}(\varepsilon, \mathscr{F}, P^m) \to 0$ *implies that* $\mathscr{F}$ *satisfies the CLT* (4.8).

PROOF. We can assume $F \le 1$ and that the functions in $\mathscr{F}$ are symmetric $(N_{[]}^{(1)}(\varepsilon, S_m\mathscr{F}, P^m) \le N_{[]}^{(1)}(\varepsilon, \mathscr{F}, P^m))$. By hypothesis $N_{[]}^{(1)}(\varepsilon n^{-1/2}, \mathscr{F}, P^m) = e^{c_n n^{1/2}/\varepsilon}$ for some $c_n \to 0$. Let $\{f_i\}_{i=1}^N$ and $\{\Delta_i\}_{i=1}^N$ be a set of functions defining $N := N_{[]}^{(1)}(\varepsilon, \mathscr{F}, P^m))$. If $|f - f_i| \le \Delta_i$ and $|g - f_i| \le \Delta_i$, then $d_{n,1}(f, g) \le 2P_n\Delta_i$. So if $\tau_n = \max\{2P_n^m(\Delta_i)\colon 1 \le i \le N\}$ we have $N_{n,1}(\tau_n, \mathscr{F}) \le e^{c_n n^{1/2}/\varepsilon}$ and therefore, letting $N_{n,1}^*$ denote the measurable envelope of $N_{n,1}$,

$$\begin{aligned}
&\Pr\left\{N_{n,1}(2\varepsilon n^{-1/2}, \mathscr{F})^* > e^{c_n n^{1/2}/\varepsilon}\right\}\\
&\quad\le \Pr^*\left\{N_{n,1}(2\varepsilon n^{-1/2}, \mathscr{F}) > e^{c_n n^{1/2}/\varepsilon}\right\}\\
&\quad\le \Pr^*\left\{N_{n,1}(2\varepsilon n^{-1/2}, \mathscr{F}) \ge N_{n,1}(\tau_n, \mathscr{F})\right\} \le \Pr\left\{\tau_n \ge 2\varepsilon n^{-1/2}\right\}\\
&\quad\le N \max_{1\le i\le N}\left[\Pr\left\{(P_n^m - P^m)(\Delta_i) > \varepsilon n^{-1/2}\right\}\right]\\
&\quad\le \exp\left\{c_n n^{1/2}/\varepsilon - 3n^{1/2}\varepsilon/14m\right\},
\end{aligned}$$

where in the last inequality we use Bernstein's inequality [Proposition 2.3(a)]. Hence, as in the last proof, we obtain

$$\begin{aligned}
&E^*\left(\log N_{n,1}(\varepsilon n^{-1/2}, \mathscr{F})\right)\\
&\quad\le \left(c_n n^{1/2}/\varepsilon\right) + n^m \log\left(n^{1/2}/\varepsilon\right)\exp\left\{c_n n^{1/2}/\varepsilon - 3n^{1/2}\varepsilon/14m\right\}
\end{aligned}$$

so that condition (4.7) holds and the result follows from Theorem 4.4. □

PROPOSITION 4.7. *If $\mathscr{F}$ is the class of indicator functions of the closed convex sets in $[0,1]^2$ and P is Lebesgue measure on $[0,1]$ (or any probability with a density f such that $0 < a \le f \le b < \infty$ for some a and b), then the CLT as in (4.8) holds for $\mathscr{F}$ and P.*

PROOF. Bronštein's result mentioned in the proof of Example 3.9 implies $\log N_{[\,]}^{(1)}(\varepsilon, \mathscr{F}, P^2) \le c\varepsilon^{-1/2}$ and in particular $\int_0^\infty (\log N(\varepsilon, P\mathscr{F}, d))^{1/2}\, d\varepsilon < \infty$. So $\mathscr{F}$ satisfies the conditions of Theorem 4.6. □

(b) *The unbounded case.* First we prove an analogue of Nolan and Pollard [(1988), Theorem 7], with slight improvements. Although this result covers VC-subgraph classes satisfying $P^m F^2 < \infty$, a better result for these classes is possible (Theorem 4.9). We also give an extension of Ossiander's (1987) CLT for empirical processes under bracketing conditions to U-processes [although an extension of the sharper result of Andersen, Giné, Ossiander and Zinn (1988) is possible, for clarity of exposition we only give Ossiander's case, which is in fact quite general].

THEOREM 4.8. *If the measurable class $\mathscr{F}$ of symmetric functions on S^m satisfies that $P^{m-1}\mathscr{F}$ is P-Donsker and that*

$$E\int_0^\infty n^{-1/2} \log N_{n,2}(\varepsilon, \mathscr{F})\, d\varepsilon \to 0, \tag{4.9}$$

then (4.1) holds for $\mathscr{F}$ and P.

PROOF. We just note that by Proposition 2.7,

$$E\left\| n^{-m+1/2} \sum_{(i_1,\dots,i_m)\in I_m^n} \varepsilon_{i_1}^{(1)}\varepsilon_{i_2}^{(2)} f(X_{i_1},\dots,X_{i_m})\right\|_{\mathscr{F}} \le cE^*\int_0^\infty n^{-1/2}\log N_{n,2}(\varepsilon,\mathscr{F})\, d\varepsilon.$$

Now, Corollary 4.2 gives the result. □

THEOREM 4.9. *Let $\mathscr{F}$ be a measurable class of functions on S^m, $P^m F^2 < \infty$ such that:*

(i) *$P^{m-1}\mathscr{F}$ is P-pre-Gaussian,*
(ii) *$t\Pr\{P^{m-1}F^2 > t\} \to 0$, and*
(iii) *there exist $c, v < \infty$ such that for all $\varepsilon > 0$ and for all probability measures Q with $QF^2 < \infty$,*

$$N(\varepsilon, \mathscr{F}, \|\cdot\|_{L_2(Q)}) \le c\left((QF^2)^{1/2}/\varepsilon\right)^v. \tag{4.10}$$

Then $\mathscr{F}$ satisfies the CLT (4.1).

PROOF. Since $Q(P^{m-1}(f-g))^2 \le QP^{m-1}(f-g)^2$, we have

$$N\big(\varepsilon, P^{m-1}\mathscr{F}, \|\cdot\|_{L_2(Q)}\big) \le N\big(\varepsilon, \mathscr{F}, \|\cdot\|_{L_2(QP^{m-1})}\big) \le c\Big[\big(QP^{m-1}F^2\big)^{1/2}/\varepsilon\Big]^v.$$

Hence, the class $P^{m-1}\mathscr{F}$ satisfies the entropy condition (iii) with respect to the envelope $(P^{m-1}F^2)^{1/2}$. Then, $P^{m-1}\mathscr{F}$ satisfies the central limit theorem by Alexander's CLT [Alexander (1987)].

Let $\alpha \in (2m/(m+1), 2)$. Condition (ii) implies that for $0 < \varepsilon < 2$, $EF^{2-\varepsilon} \le P(P^{m-1}F^2)^{(2-\varepsilon)/2} < \infty$, therefore, $n^m P^m(F > n^{m/\alpha}) \to 0$. So, since

$$E_\varepsilon\bigg|n^{1/2-m}\sum_{I_m^n}\varepsilon_{i_1}\varepsilon_{i_2} f\big(X_{i_1},\dots,X_{i_m}\big)\bigg|^2 \le n^{-m-1}\sum_{I_m^n} f^2\big(X_{i_1},\dots,X_{i_m}\big),$$

we have

$$\begin{aligned}
&E\bigg\|n^{1/2-m}\sum_{I_m^n}\varepsilon_{i_1}\varepsilon_{i_2} f\big(X_{i_1},\dots,X_{i_m}\big)I_{F\le n^{m/\alpha}}\bigg\|_{\mathscr{F}}\\
&\quad\le\Bigg[E\bigg(n^{-m-1}\sum_{I_m^n}F^2 I_{F\le n^{m/\alpha}}\big(X_{i_1},\dots,X_{i_m}\big)\bigg)^2\Bigg]^{1/2} \le \big(n^{-1}EF^2 I_{F\le n^{m/\alpha}}\big)^{1/2}\\
&\quad\le\bigg[2n^{-1}\int_0^{n^{m/\alpha}} tP\{F>t\}\,dt\bigg]^{1/2} \le \bigg(n^{-1}\int_0^{n^{m/\alpha}} ct^{1-\alpha}\,dt\bigg)^{1/2}\\
&\quad\le \big(n^{(2-\alpha)m/\alpha-1}\big)^{1/2} \to 0.
\end{aligned}$$

This, together with the condition $n^m P^m(F > n^{m/\alpha}) \to 0$, implies, by the second part of the proof of Corollary 4.2, that $n^{1/2}\|U_m^n(\pi_{k,m} f)\|_{\mathscr{F}} \to 0$ in probability. Hence, (b) in Corollary 4.2 is satisfied and the result follows. □

Theorem 4.9 applies to VC-subgraph classes of functions by Lemma II.25 in Pollard (1984) and Lemma 4.4 in Alexander (1987). This case, under stronger integrability hypotheses, has been considered before: Schneemeier (1989) proves the CLT for VC classes of sets in the triangular arrays case, and Sherman (1991) obtains several results related to Theorem 4.9, and gives some interesting statistical applications (his work and ours are independent and approximately simultaneous).

The inequality in Proposition 2.3(a) shows that if $Ef = 0$, $Ef^2 = \sigma^2$ and $\|f\|_\infty \le c$, then $P\{n^{1/2}|U_m^n(f)| > t\} \le e^{-t^2/3m\sigma^2}$, for $t \lesssim \sigma'^2 n^{1/2}/c$, exactly the same breakpoint as for sums of i.i.d. bounded random variables. A consequence of this last fact is that the bracketing CLT for nondegenerate U-processes can be proved in essentially the same way as for empirical processes, at least in Ossiander's (1987) version.

THEOREM 4.10. *Let $\mathscr{F}$ be a class of functions on S^m. If*

$$\int_0^\infty\big(\log N_{[\,]}^{(2)}(\varepsilon, \mathscr{F}, P^m)\big)^{1/2}\,d\varepsilon < \infty,$$

then

$$\mathscr{L}\big(n^{1/2}(U_m^n - P^m)f\big) \to_w \mathscr{L}\big(mG_P \circ P^{m-1}f\big) \quad in\ l^\infty(\mathscr{F}).$$

PROOF. Note that the bracketing condition implies $PF^2 < \infty$. We may assume, as in Theorem 4.4, that $\mathscr{F}$ is a class of symmetric functions. Let $N_k = N_{[\,]}^{(2)}(2^{-k})$. Then by hypothesis, $\sum_q 2^{-1}(\log qN_1 \cdots N_q)^{1/2} < \infty$. Let $\gamma_q = (\log qN_1 \cdots N_q)^{1/2}$. If $\beta_{q_0} = \sum_{q=q_0}^\infty 2^{-q}\gamma_q$, we have $\beta_{q_0} \to 0$ as $q_0 \to \infty$. Also, $\gamma_q \nearrow$ as $q \nearrow$. Let $\bar{A}_{q,1}, \dots, \bar{A}_{q,N_q}$ be an optimal set of 2^{-q} brackets (that we take to be a partition of $\mathscr{F}$). Define for $i_j \le N_j$, $j \le q$, $A_{q,i_1,\dots,i_q} = \bigcap_{j=1}^q \bar{A}_{j,i_j}$. Then, obviously $\#\{A_{q,i_1,\dots,i_q}: i_j \le N_j\} \le N_1 \cdots N_q < e^{\gamma_q^2}$. Fix a function $f_{q,i_1,\dots,i_q}$ in each $A_{q,i_1,\dots,i_q}$. Relabel $A_q(f) = A_{q,i_1,\dots,i_q}$, $\pi_q(f) = f_{q,i_1,\dots,i_q}$ if $f \in A_{q,i_1,\dots,i_q}$, and define $\Delta_q(f) = \sup_{g \in A_q(f)}|g - \pi_q f|$. Then we have $\Delta_q(f) \searrow$ as $q \nearrow$ for every $f \in \mathscr{F}$ and $E\Delta_q^2(f) \le 2^{-2q}$. Finally, define, for any given q_0,

$$\tau f = \min\big\{q \ge q_0: \Delta_q f > n^{1/2}2^{-q-1}\gamma_{q+1}^{-1}\big\},$$

with $\min \varnothing = \infty$. Obviously,

$$\{\tau f > q_0\} = \big\{\Delta_{q_0} f \le n^{1/2}2^{-q_0-1}\gamma_{q_0+1}^{-1}\big\},$$

$$\{\tau f = q_0\} = \big\{\Delta_{q_0} f > n^{1/2}2^{-q_0-1}\gamma_{q_0+1}^{-1}\big\},$$

$$\{\tau f \ge q\} \subset \big\{\Delta_{q-1} f \le n^{1/2}2^{-q}\gamma_q^{-1}\big\} \subset \big\{\Delta_q f \le n^{1/2}2^{-q}\gamma_q^{-1}\big\}$$

and $\{\tau f = q\} \subset \{n^{1/2}2^{-q-1}\gamma_{q+1}^{-1} < \Delta_q f \le n^{1/2}2^{-q}\gamma_q^{-1}\}$, where we are using that $\Delta_q f$ decreases.

We must prove

$$\lim_{q_0\to\infty} \limsup_{n\to\infty} \Pr{}^*\big\{n^{1/2}\|(U_m^n - P^m)(f - \pi_{q_0} f)\|_{\mathscr{F}} > \varepsilon\big\} = 0 \tag{4.11}$$

for all $\varepsilon > 0$.

For this we decompose $f - \pi_{q_0} f$ as follows [as in Andersen, Giné, Ossiander and Zinn (1988)]:

$$f - \pi_{q_0} f = (f - \pi_{q_0} f)I_{\tau f = q_0} + (f - \pi_{q_1} f)I_{\tau f \ge q_1} + \sum_{q=q_0+1}^{q_1-1}(f - \pi_q f)I_{\tau f = q} + \sum_{q=q_0+1}^{q_1}(\pi_q f - \pi_{q-1} f)I_{\tau f \ge q}.$$

This decomposition of $f - \pi_{q_0} f$ induces a decomposition of the probability in (4.11) into four parts that we label (I), (II), (III) and (IV).

(I) Taking q_0 such that $\beta_{q_0} < \varepsilon/24m$ we have

$$n^{1/2} E\Delta_{q_0} f I_{\Delta_{q_0} f > 2^{-q_0-1} n^{1/2} \gamma_{q_0+1}^{-1}} \le 2^{q_0+1} \gamma_{q_0+1} E(\Delta_{q_0} f)^2$$
$$\le 2^{-q_0+1} \gamma_{q_0+1} \le 4\beta_{q_0} < \varepsilon/4$$

and therefore

$$\Pr{}^*\left\{\left\| n^{1/2}(U_m^n - P^m)(f - \pi_{q_0} f) I_{\tau f = q_0}\right\|_{\mathscr{F}} > 2\varepsilon\right\}$$
$$\le \Pr\left\{\left\| n^{1/2}(U_m^n - P^m)\left(\Delta_{q_0} f I_{\Delta_{q_0} f > 2^{-q_0-1} n^{1/2} \gamma_{q_0+1}^{-1}}\right)\right\|_{\mathscr{F}} > \varepsilon\right\}$$
$$\le e^{\gamma_{q_0}^2} \max_{\Delta_{q_0} f} \Pr\left\{ n^{1/2}\left|(U_m^n - P^m)\left(\Delta_{q_0} f I_{\Delta_{q_0} f > 2^{-q_0-1} n^{1/2} \gamma_{q_0+1}^{-1}}\right)\right| > \varepsilon\right\}$$
$$\le e^{\gamma_{q_0}^2} \varepsilon^{-2} m \max_{\Delta_{q_0} f} \operatorname{Var}\left(\Delta_{q_0} f I_{\Delta_{q_0} f > 2^{-q_0-1} n^{1/2} \gamma_{q_0+1}^{-1}}\right) \to 0,$$

as $n \to \infty$ for all q_0.

(II) Taking q_1 such that $n^{1/2} 2^{-q_1} < \varepsilon/4$ we have that

$$n^{1/2}\left| E\Delta_{q_1} f I_{\Delta_{q_1} f \le n^{1/2} 2^{-q_1} \gamma_{q_1}^{-1}}\right| \le n^{1/2}\left(E\Delta_{q_1} f^2\right)^{1/2} \le n^{1/2} 2^{-q_1} < \varepsilon/4$$

and

$$E\Delta_{q_1}^2 f I_{\Delta_{q_1} f \le n^{1/2} 2^{-q_1} \gamma_{q_1}^{-1}} \le \varepsilon 2^{-q_1-2} \gamma_{q_1}^{-1}.$$

Now we can replace $f - \pi_{q_1} f$ by $\Delta_{q_1} f$ in

$$\Pr\left\{\| n^{1/2}(U_m^n - P^m)\left((f - \pi_{q_1} f) I_{\tau f = q_1}\right)\|_{\mathscr{F}} > 2\varepsilon\right\},$$

and center, as in (I), and then apply Bernstein's inequality [Proposition 2.3(a)] to obtain

$$\Pr{}^*\left\{\left\| n^{1/2}(U_m^n - P^m)(f - \pi_{q_1} f) I_{\tau f = q_1}\right\|_{\mathscr{F}} > 2\varepsilon\right\}$$
$$\le \Pr\left\{\left\| n^{1/2}(U_m^n - P^m)\left(\Delta_{q_1} f I_{\Delta_{q_1} f < n^{1/2} 2^{-q_1-1} \gamma_{q_1}^{-1}}\right)\right\|_{\mathscr{F}} > \varepsilon\right\}$$
$$\le \exp\left\{\gamma_{q_1}^2 - \varepsilon^2 / \left(2m\left[\varepsilon 2^{-q_1-1} \gamma_{q_1}^{-1} + (2/3)\varepsilon 2^{-q_1} \gamma_{q_1}^{-1}\right]\right)\right\}$$
$$\le \exp\left\{\gamma_{q_1}^2 - (3/7m)\varepsilon 2^{q_1} \gamma_{q_1}\right\}$$
$$\le \exp\left\{-m^{-1} \varepsilon 2^{q_1-2} \gamma_{q_1}\right\}$$

since $2^{-q_1} \gamma_{q_1} \le \beta_{q_0} < \varepsilon/24m$.

(III) Since $n^{1/2}E\Delta_q fI_{\tau f=q} \le 2^q\gamma_q E(\Delta_q f)^2 \le 2^{-q}\gamma_q$ and since $\beta_{q_0} \to 0$, we have, using Bernstein's inequality once more,

$$\begin{aligned}
\Pr{}^*&\left\{\left\| n^{1/2}(U_m^n - P^m)\sum_{q=q_0+1}^{q_1-1}(f-\pi_q f)I_{\tau f=q}\right\|_{\mathscr{F}} > 2\varepsilon\right\} \\
&\le \sum_{q=q_0+1}^{q_1-1} \Pr\left\{\left\| n^{1/2}(U_m^n - P^m)(f-\pi_q f)I_{\tau f=q}\right\|_{\mathscr{F}} > \varepsilon 2^{-q}\gamma_{q+1}\beta_{q_0}^{-1}\right\} \\
&\le \sum_{q=q_0+1}^{q_1-1} \Pr\left\{\left\| n^{1/2}(U_m^n - P^m)(\Delta_q fI_{\tau f=q})\right\|_{\mathscr{F}} > \varepsilon 2^{-q-1}\gamma_{q+1}\beta_{q_0}^{-1}\right\} \\
&\le \sum_{q=q_0+1}^{q_1-1} \exp\left\{\gamma_q^2 - \varepsilon^2 2^{-2q-2}\gamma_q^2\beta_{q_0}^{-2}/\left(2m\left[2^{-2q+1} + (2/3)\varepsilon 2^{-2q-1}\beta_{q_0}^{-1}\right]\right)\right\} \\
&\le \sum_{q=q_0+1}^{q_1-1} \exp\left\{\gamma_q^2 - 2^{-2}m^{-1}\varepsilon\gamma_q^2\beta_{q_0}^{-1}\right\} \le \sum_{q=q_0+1}^{q_1-1} \exp\left\{-2^{-3}m^{-1}\varepsilon\gamma_q^2\beta_{q_0}^{-1}\right\}.
\end{aligned}$$

(IV) We just recall that $g \in A_q(f)$ implies $\pi_q(g) = \pi_q(f)$, $\pi_{q-1}(g) = \pi_{q-1}(f)$ and $\{\tau f = q\} = \{\tau g = q\}$ because $\Delta_r f = \Delta_r g$ for all $r \le q$. So, $\#\{(\pi_q f - \pi_{q-1} f)I_{\tau f \ge q}\} \le e^{\gamma_q^2}$. Also, $E(\pi_q f - \pi_{q-1} f)^2 \le 2^{-2(q-1)}$ because $\pi_q f \in A_{q-1}(f)$. We then have, by Bernstein's inequality,

$$\begin{aligned}
\Pr{}^*&\left\{\left\| n^{1/2}\sum_{q=q_0+1}^{q_1}(U_m^n - P^m)(\pi_q f - \pi_{q-1} f)I_{\tau f\ge q}\right\|_{\mathscr{F}} > \varepsilon\right\} \\
&\le \sum_{q=q_0+1}^{q_1} \Pr\left\{\left\| n^{1/2}(U_m^n - P^m)(\pi_q f - \pi_{q-1} f)I_{\tau f\ge q}\right\|_{\mathscr{F}} > \varepsilon 2^{-q}\gamma_q\beta_{q_0}^{-1}\right\} \\
&\le \sum_{q=q_0+1}^{q_1} \exp\left\{\gamma_q^2 - \varepsilon^2 2^{-2q}\gamma_q^2\beta_{q_0}^{-2}\Big/\left(2m\left[2^{-2q+3} + (2/3)\varepsilon 2^{-2q}\beta_{q_0}^{-1}\right]\right)\right\} \\
&\le \sum_{q=q_0+1}^{q_1} \exp\left\{\gamma_q^2 - (2m)^{-1}\varepsilon\beta_{q_0}^{-1}\gamma_q^2\right\} \\
&\le \sum_{q=q_0+1}^{q_1} \exp\left\{-(1/4m)\varepsilon\beta_{q_0}^{-1}\gamma_q^2\right\} \quad \text{if } \beta_{q_0} < \varepsilon/24m.
\end{aligned}$$

CONCLUSION.

$$\begin{aligned}
\lim_{q_0\to\infty} \limsup_n &((\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}) + (\mathrm{IV})) \\
&\le \lim_{q_0\to\infty} \sum_{q=q_0+1}^{\infty} 2\exp\left\{-(\varepsilon/8m\beta_{q_0})\gamma_q^2\right\} \\
&\le \lim_{q_0\to\infty} \sum_{q=q_0+1}^{\infty} 2q^{-3} = 0,
\end{aligned}$$

since $\gamma_q^2 \ge \log q$ and $\beta_{q_0} < \varepsilon/24m$. □

The preceding proof, adapted from Andersen, Giné, Ossiander and Zinn (1988), provides, for $m = 1$, a streamlined approach to Ossiander's (1987) theorem.

5. The CLT for canonical classes of functions. A blanket assumption on classes $\mathscr{F}$ of canonical functions $f\colon S^m \to \mathbb{R}$ in this section will be

$$\int_0^\infty \left(\log N(\varepsilon, \mathscr{F}, e_{P,m})\right)^{m/2} d\varepsilon < \infty. \tag{5.1}$$

It ensures that the paths of a version of the limit process $\{K_{P,m}(f)\colon f \in \mathscr{F}\}$ are bounded and uniformly continuous on $(\mathscr{F}, e_{P,m})$ by Proposition 2.7. In this section $\mathscr{F}_\delta'$ will denote the set of functions $\{f - g\colon f, g \in \mathscr{F},\ e_{P,m}(f,g) \le \delta\}$.

(a) *The case of uniformly bounded* $\mathscr{F}$. If $\mathscr{F}$ is uniformly bounded, then condition (5.1) allows us to weaken the asymptotic equicontinuity condition (1.7), just as in Giné and Zinn [(1984), Theorem 3.1]. Instead of Bernstein's inequality one uses here inequality (c) in Proposition 2.3, and the proof consists of a partial chaining up to the level where the bound in this inequality stops being of the order $\exp(-(t/\sigma)^{2/m})$, in complete analogy with the empirical process case. The proof is omitted.

THEOREM 5.1. *Let $\mathscr{F}$ be a measurable uniformly bounded class of canonical functions $f\colon S^m \to \mathbb{R}$ satisfying* (5.1) *and*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \Pr\left\{ \sup_{e_{P,m}(f,g) \le \delta n^{-m/2(m+1)}} \left| n^{-m/2} \times \sum_{(i_1,\dots,i_m)\in I_k^n} (f-g)(X_{i_1},\dots,X_{i_m}) \right| > \varepsilon \right\} = 0 \tag{5.2}$$

for all $\varepsilon > 0$. Then $\mathscr{F}$ satisfies the CLT, that is,

$$n^{m/2} U_m^n \to_{\mathscr{L}} K_{P,m} \circ S_m(f) \quad \text{in } l^\infty(\mathscr{F}). \tag{5.3}$$

With Theorem 5.1 we can obtain sufficient conditions for the CLT in terms of random entropies.

THEOREM 5.2. *Let $\mathscr{F}$ be a measurable uniformly bounded class of symmetric canonical functions satisfying the entropy condition* (5.1). *If*
(a) *$\int_0^\infty (\log N_{n,2}(\varepsilon, \mathscr{F}))^{m/2}\, d\varepsilon$ is uniformly integrable and*

(b) $$\lim_{\delta\to 0} \limsup_{n\to\infty} E^* \int_0^{\delta n^{-m/2(m+1)}} \left(\log N_{n,2}(\varepsilon, \mathscr{F})\right)^{m/2} d\varepsilon = 0,$$

then (5.3) *holds.*

PROOF. Set $\mathscr{F}'_{\delta,n} = \{f - g: \ f, g \in \mathscr{F}, e_{P,m}(f, g) \le \delta n^{-m/2(m+1)}\}$. By Propositions 2.1 and 2.6 we have

$$\begin{aligned}
E\|n^{m/2}U_m^n(f)\|_{\mathscr{F}'_{\delta,n}} &\lesssim E\Big\|n^{-m/2}\sum_{I_m^n}\varepsilon_{i_1}\cdots\varepsilon_{i_m}f(X_{i_1},\dots,X_{i_m})\Big\|_{\mathscr{F}'_{\delta,n}}\\
&\lesssim E^*\left(\int_0^\infty(\log N_{n,2}(\varepsilon,\mathscr{F}))^{m/2}\,d\varepsilon\right)\\
&\lesssim E^*\left(\int_0^{5^{1/2}\delta n^{-m/2(m+1)}}(\log N_{n,2}(\varepsilon,\mathscr{F}))^{m/2}\,d\varepsilon\right)\\
&\qquad+E^*\left(\int_0^\infty(\log N_{n,2}(\varepsilon,\mathscr{F}))^{m/2}\,d\varepsilon\, I_{D_n^2(\delta)>5\delta^2n^{-m/(m+1)}}\right),
\end{aligned}$$

where

$$\begin{aligned}
D_n^2(\delta) &= \sup_{f,g\in\mathscr{F}_\delta'}\left[n^{-m}\sum_{(i_1,\dots,i_m)\in I_m^n}(f-g)^2(X_{i_1},\dots,X_{i_m})\right]\\
&\le 4\Big\|n^{-m}\sum_{(i_1,\dots,i_m)\in I_m^n}f^2(X_{i_1},\dots,X_{i_m})\Big\|_{\mathscr{F}'_{\delta,n}}.
\end{aligned}$$

Since $e_{n,1}(f^2,g^2)\le 2e_{n,2}(f,g)$, hypothesis (a) gives convergence to 0 of $E^*(n^{-1}\log N_{n,1}(\varepsilon,(\mathscr{F}')^2))$. Hence the law of large numbers (Corollary 3.2) applied to $(\mathscr{F}')^2$ gives $\Pr\{D_n^2(\delta)>5\delta^2n^{-m/(m+1)}\}\to 0$ for all $\delta>0$. So, the hypotheses imply that the equicontinuity condition (5.2) is satisfied and Theorem 5.1 applies. □

Next we restrict our attention to the case $m = 2$, which admits a better development because of the following proposition. (Here $\mathscr{F}'_{\delta,n}$ is as defined in the proof of Theorem 5.2.)

PROPOSITION 5.3. *Let $\mathscr{F}$ be a measurable class of symmetric canonical uniformly bounded functions on S^2, and let P be such that $\mathscr{F}$ and P satisfy condition* (5.1). *Then the following are equivalent*:

(a) *$\mathscr{F}$ satisfies the CLT* (5.3).
(b) $\lim_{\delta\to\infty}\limsup_{n\to\infty}\Pr\{\|nU_2^n(f,P)\|_{\mathscr{F}'_{\delta,n}}>\varepsilon\} = 0$ *for all* $\varepsilon>0$.
(c) $\lim_{\delta\to\infty}\limsup_{n\to\infty}E\|nU_2^n(f,P)\|^r_{\mathscr{F}'_{\delta,n}} = 0$ *for all* (*some*) $r<\infty$.
(d) $\lim_{\delta\to\infty}\limsup_{n\to\infty}\Pr\{\|n^{-1}\Sigma_{I_2^n}\varepsilon_i\varepsilon_j f(X_i,X_j)\|_{\mathscr{F}'_{\delta,n}}>\varepsilon\} = 0$ *for all* $\varepsilon>0$.
(e) $\lim_{\delta\to\infty}\limsup_{n\to\infty}E\|n^{-1}\Sigma_{I_2^n}\varepsilon_i\varepsilon_j f(X_i,X_j)\|^r_{\mathscr{F}'_{\delta,n}} = 0$ *for all* (*some*) $r<\infty$.
(f) *Any of the decoupled versions of* (d) *and* (e).

PROOF. By Proposition 2.8,

$$E\left\|n^{-1}\sum_{I_2^n}\varepsilon_i\varepsilon'_j f(X_i, X'_j)\right\|^p \le E\max_{i\le n}\left\|n^{-1}\sum_{I_2^n}\varepsilon'_j f(X_i, X'_j)\right\|^p + c\left(E\left\|n^{-1}\sum_{I_2^n}\varepsilon_i\varepsilon'_j f(X_i, X'_j)\right\|\right)^p.$$

By Proposition 2.1, these same inequalities hold for $n^{-1}\Sigma_{I_2^n}\varepsilon_i\varepsilon_j f(X_i, X_j)$ and $nU_2^n(f, P)$. Now, by Theorem 5.1, Paley–Zygmund's inequality [Kahane (1968), page 6] gives all the equivalences. □

Proposition 2.8 does not seem to provide enough uniform integrability to deduce the analogue of Proposition 5.3 for $m > 2$. It would be surprising, however, if 5.3 did not extend to $m > 2$.

THEOREM 5.4. *Let $\mathscr{F}$ be a measurable uniformly bounded class of symmetric canonical functions on S^2 satisfying the entropy condition* (5.1).

(a) *Under the hypotheses* (i) $n^{-1/3}\log N_{n,1}^{\mathrm{dec}}(\varepsilon n^{-1}, \mathscr{F}) \to 0$ *in* pr* *for all* $\varepsilon > 0$, *and* (ii) $n^{-1/3}\log N(\varepsilon n^{-1/3}, \mathscr{F}, \|\cdot\|_{L_2(P_n\times P)}) \to 0$ *in probability* for all* $\varepsilon > 0$, $\mathscr{F}$ *satisfies the CLT* (5.3).

(b) *If*

$$\lim_{\delta\to 0}\limsup_{n\to\infty} E^*\int_0^{\delta n^{-1/3}}\log N_{n,2}(\varepsilon, \mathscr{F})\, d\varepsilon = 0,$$

then $\mathscr{F}$ satisfies the CLT (5.3).

PROOF OF THEOREM 5.4(a). It suffices to show

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr\left\{n^{-1}\left\|\sum_{(i,j)\in I_2^n}\varepsilon_i\varepsilon'_j f(X_i, X'_j)\right\|_{\mathscr{F}'_{\delta,n}} > \varepsilon\right\} = 0 \tag{5.2'}$$

for all $\varepsilon > 0$. We decompose these probabilities as follows:

$$\begin{aligned}
&\Pr\left\{\left\|\sum_{(i,j)\in I_2^n}\varepsilon_i\varepsilon'_j f(X_i, X'_j)\right\|_{\mathscr{F}'_{\delta,n}} > 4\varepsilon n\right\}\\
&\quad\le \Pr\left\{n^{-1/3}\log N_{n,1}^{\mathrm{dec}}(\varepsilon n^{-1}, \mathscr{F}) > 1\right\}\\
&\qquad + \Pr\left\{\left\|\sum_{(i,j)\in I_2^n} f^2(X_i, X'_j)\right\|_{\mathscr{F}'_{\delta,n}} > 256\delta^2 n^{4/3}\right\}\\
&\qquad + \Pr\left\{\left\|\sum_{(i,j)\in I_2^n}\varepsilon_i\varepsilon'_j f(X_i, X'_j)\right\|_{\mathscr{F}'_{\delta,n}} > 4\varepsilon n, \log N_{n,1}^{\mathrm{dec}}(\varepsilon n^{-1}, \mathscr{F}) \le n^{1/3},\right.\\
&\qquad\qquad\left.\left\|\sum_{(i,j)\in I_2^n} f^2(X_i, X'_j)\right\|_{\mathscr{F}'_{\delta,n}} \le 256\delta^2 n^{4/3}\right\}\\
&\quad =: (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}).
\end{aligned}$$

By Proposition 2.2, $\Pr\{\Sigma_{(i,j)\in I_2^n}\varepsilon_i\varepsilon_j' a_{ij} \ge t\} \le \exp(2 - e^{-1}t(\Sigma a_{ij}^2)^{-1/2})$. We can apply this bound to (III) conditionally on $\{X_i, X_j'\}$ and obtain (III) $\le \exp(2n^{1/3} + 2 - (\varepsilon n^{1/3}/8\delta e))$. [Note $N_{n,1}^{\mathrm{dec}}(2\varepsilon n^{-1}, \mathscr{F}_{\delta,n}') \le (N_{n,1}^{\mathrm{dec}}(\varepsilon n^{-1}, \mathscr{F}))^2$.] Hence $\lim_{\delta\to 0}\limsup_{n\to\infty}$ (III) $= 0$ for all $\varepsilon > 0$.

By hypothesis (i) we also have $\lim_{n\to\infty}$(I) $= 0$ for all $\varepsilon > 0$.

In order to estimate (II) we twice apply the version of Le Cam's "square root trick" inequality in Lemma 5.2 in Giné and Zinn (1984). First we need to symmetrize. Since $\sup_{\mathscr{F}_{\delta,n}'}\Pr\{\Sigma_{i,j=1}^n f^2(X_i, X_j') > 128\delta^2 n^{4/3}\} < 1/2$, the symmetrization Lemma 2.5 in Giné and Zinn (1984) gives (II) $\le \Pr\{\|\Sigma_{i,j=1}^n(f^2(X_i, X_j') - f^2(Y_i, Y_j'))\|_{\mathscr{F}_{\delta,n}'} > 256\delta^2 n^{4/3}\}$, where $\{Y_i, Y_j'\}$ is an independent copy of $\{X_i, X_j'\}$. By adding and subtacting $\Sigma f^2(X_i, Y_j')$ the sums become conditionally symmetric and Rademacher randomization can be introduced. Assuming as we can that the functions f are symmetric, we conclude

$$\text{(II)} \le 8\Pr\left\{\left\|\sum_{i=1}^n\sum_{j=1}^n \varepsilon_i f^2(X_i, X_j')\right\|_{\mathscr{F}_{\delta,n}'} > 32\delta^2 n^{4/3}\right\}.$$

Let E_X and $E_{X'}$ denote conditional expectation with respect to $\{X_i\}$ and $\{X_i'\}$, respectively. Thus

$$\begin{aligned}\text{(II)} \le 8\Pr\Bigg\{&\left\|\sum_{i=1}^n\sum_{j=1}^n \varepsilon_i f^2(X_i, X_j')\right\|_{\mathscr{F}_{\delta,n}'} \ge 32\delta^2 n^{4/3},\\ &\left\|\sum_{j=1}^n E_X f^2(X_1, X_j')\right\|_{\mathscr{F}_{\delta,n}'} \le 6\delta^2 n^{1/3}\Bigg\}\\ &+ 8\Pr\left\{\left\|\sum_{j=1}^n E_X f^2(X_1, X_j')\right\|_{\mathscr{F}_{\delta,n}'} > 6\delta^2 n^{1/3}\right\} = \text{(II)}_1 + \text{(II)}_2.\end{aligned}$$

We can apply Lemma 5.2 in the cited reference to (II)_1 conditionally on $\{X_j'\}$, with $t = 32\delta^2 n^{5/6}$, $M_n = 6\delta^2 n^{5/6}$, $m = \exp(2^{-6}\delta^2 n^{1/3})$, $\rho = 2^{-3}\delta n^{5/12}$, $\lambda = \delta 2^{-2} n^{5/12}$ and $r = n$ (we assume $F \le 1/2$) to get

$$\begin{aligned}\text{(II)}_1 \le \Pr{}^*&\left\{N_{n,2}(2^{-3}\delta n^{1/6}), \left\{\sum_j f^2(X_i, X_j') : f \in \mathscr{F}_{\delta,n}'\right\}\right\} > \exp(2^{-6}\delta^2 n^{1/3})\Bigg\}\\ &+ 64\exp(-2^{-6}\delta^2 n^{1/3}).\end{aligned}$$

Because of the inequality

$$\begin{aligned}&\left(n^{-1}\sum_{i=1}^n\left[\left(\sum_{j=1}^n f^2(X_i, X_j')\right)^{1/2} - \left(\sum_{j=1}^n g^2(X_i, X_j')\right)^{1/2}\right]^2\right)^{1/2}\\ &\qquad\le \left(n^{-1}\sum_{i,j=1}^n (f-g)^2(X_i, X_j')\right)^{1/2} \le \left(n e_{n,1}^{\mathrm{dec}}(f, g) + 1\right)^{1/2},\end{aligned}$$

the above covering number is dominated as follows:

$$N_{n,1}^{\mathrm{dec}}(2^{-6}\delta^2 n^{-2/3} - n^{-1}, \mathscr{F}_{\delta,n}') \le N_{n,1}^{\mathrm{dec}}(2^{-7}\delta^2 n^{-2/3}, \mathscr{F}_{\delta,n}')$$
$$\le \left(N_{n,1}^{\mathrm{dec}}(2^{-8}\delta^2 n^{-2/3}, \mathscr{F})\right)^2.$$

So $\lim_{\delta\to 0}\limsup_{n\to\infty}(\mathrm{II})_1 = 0$ by condition (i). Now we apply the square root trick to bound $(\mathrm{II})_2$. In this case we take $t = 6\delta^2 n^{-1/6}$, $M_n = \delta^2 n^{-1/6}$, $m = \exp(2^{-6}\delta^2 n^{1/3})$, $\rho = 2^{-2}\delta n^{-1/12}$, $\lambda = 2^{-2}\delta n^{-1/12}$ and $r = 1$, and, since

$$\left(n^{-1}\sum_{j=1}^{n}\left[\left(E_X f^2(X_1, X_j')\right)^{1/2} - \left(E_X g^2(X_1, X_j')\right)^{1/2}\right]^2\right)^{1/2}$$
$$\le \left(n^{-1}\sum_{j=1}^{n} E_X(f-g)^2(X_1, X_j')\right)^{1/2}$$
$$= \left[P_n \times P(f-g)^2\right]^{1/2},$$

condition (ii) and the cited lemma imply the result.

PROOF OF THEOREM 5.4(b). This follows as in Theorem 5.2 by computing probabilities instead of expected values, which is possible by Proposition 5.3. □

We define the following bracketing numbers which are appropriate for the canonical case:

$$N_{[\,],c}^{(p)}(\varepsilon, \mathscr{F}) = \min\Big\{r\colon \text{there exist } f_1, \dots, f_r \text{ and } \Delta_1, \dots, \Delta_r \in L_p(P^2) \text{ such that } (P^m\Delta_i^p)^{1/p} < \varepsilon$$
$$\text{and such that for all } f \in \mathscr{F} \text{ there exists } i \le r \text{ with } |f_i - f| < \Delta_i \text{ and } \Delta_i - P^m\Delta_i \text{ are } P\text{-canonical}\Big\}.$$

This definition is slightly weaker than the usual one with the extra requirement that the brackets $[f_i, f_j]$ be defined by functions f_i, f_j from the class $\mathscr{F}$. From Theorem 5.3, using the method of proof of Theorem 4.6, and Proposition 2.3(c) instead of Bernstein's inequality, it is possible to obtian the following corollary.

COROLLARY 5.5. *Let $\mathscr{F}$ be a measurable uniformly bounded class of canonical functions. If* (5.1) *holds and*

$$\varepsilon^{1/3}\log N_{[\,],c}^{(1)}(\varepsilon, \mathscr{F}) \to 0, \tag{5.4}$$

then

$$\Big\{n^{-1}\sum_{(i,j)\in I_2^n} f(X_i, X_j)\Big\} \to_{\mathscr{L}} \{K_{P,2}(f)\} \quad \text{in } l_\infty(\mathscr{F}).$$

(b) *The unbounded case.* We prove a CLT under random entropy conditions [see Section 8 in Giné and Zinn (1984) for an analogue in the empirical process case and Nolan and Pollard (1988) for a related result for U-processes].

THEOREM 5.6. *Let $\mathscr{F}$ be a measurable family of real P-canonical functions on S^m such that $(\mathscr{F}_\delta')^2$ is also measurable for all $\delta > 0$, satisfying*:

(a) $P^m F^2 < \infty$.
(b) *The sequence* $\{\int_0^\infty (\log N_{n,2}(\varepsilon, \mathscr{F}))^{m/2}\, d\varepsilon\}_{n=1}^\infty$ *is uniformly integrable, and*
(c) $\lim_{\delta \to 0} \limsup_{n \to \infty} E^*(\int_0^\delta (\log N_{n,2}(\varepsilon, \mathscr{F}))^{m/2}\, d\varepsilon) = 0$.

Then

$$\{n^{m/2} U_m^n f : f \in \mathscr{F}\} \to_{\mathscr{L}} \{K_{p,m}(f) : f \in \mathscr{F}\} \quad \text{in } l_\infty(\mathscr{F}).$$

PROOF. By the law of large numbers we have $N(\varepsilon, \mathscr{F}, e_{P,m}) \le \liminf N_{n,2}(\varepsilon/2, \mathscr{F})$. Hence Fatou's lemma gives

$$\begin{aligned}(\log N(\varepsilon, \mathscr{F}, e_{P,m}))^{m/2} &\le \liminf E^*(\log N_{n,2}(\varepsilon/2, \mathscr{F}))^{m/2} \\ &\le \limsup_{n\to\infty} (2/\varepsilon) E^* \int_0^{\varepsilon/2} (\log N_{n,2}(\tau/2, \mathscr{F}))^{m/2}\, d\tau,\end{aligned}$$

which is finite by condition (c). Hence $(\mathscr{F}, e_{P,m})$ is totally bounded and we need only prove the asymptotic equicontinuity condition. We have, as in Theorem 5.2,

$$E\|n^{m/2} U_m^n(f)\|_{\mathscr{F}_\delta'} \le E^* \int_0^\infty (\log N_{n,2}(\tau, \mathscr{F}))^{m/2}\, d\tau. \tag{5.5}$$

Let, as before,

$$\begin{aligned}D_n^2(\delta) &= \sup_{f,g \in \mathscr{F}_\delta'} \left[n^{-m} \sum_{(i_1,\dots,i_m)\in I_m^n} (f-g)^2(X_{i_1},\dots,X_{i_m}) \right] \\ &\le 4 \left\| n^{-m} \sum_{(i_1,\dots,i_m)\in I_m^n} f^2(X_{i_1},\dots,X_{i_m}) \right\|_{\mathscr{F}_\delta'}.\end{aligned} \tag{5.6}$$

For $f, g \in \mathscr{F}_\delta'$, $e_{n,1}(f^2, g^2) \le 2 e_{n,2}(f,g)(n^{-m}\sum_{I_m^n} F^2(X_{i_1},\dots,X_{i_m}))^{1/2}$ by Schwarz's inequality. By (a) the last factor converges a.s. to $(EF^2)^{1/2} < \infty$. Then

$$\begin{aligned}&\Pr{}^*\left\{\varepsilon\left(\log N_{n,1}(\varepsilon, (\mathscr{F}_\delta')^2)\right)^{m/2} > \tau\right\} \\ &\quad \le \Pr\left\{n^{-m} \sum_{(i_1,\dots,i_m)\in I_m^n} F^2(X_{i_1},\dots,X_{i_m}) > 4EF^2\right\} \\ &\qquad + \Pr{}^*\left\{\varepsilon\left(\log N_{n,2}(\varepsilon/4(EF^2)^{1/2}, \mathscr{F}_\delta')\right)^{m/2} > \tau\right\}.\end{aligned} \tag{5.7}$$

Note first that $\log N_{n,2}(\varepsilon, \mathscr{F}_\delta') \le 2\log N_{n,2}(2^{-1}\varepsilon, \mathscr{F})$ and then apply condition (c) together with inequality (5.7) to get

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty} \Pr\Big\{\varepsilon\Big(\log N_{n,1}\big(\varepsilon, (\mathscr{F}_\delta')^2\big)\Big)^{m/2} > \tau\Big\} = 0$$

for all $\tau, \delta > 0$. Hence $\lim_{n\to\infty}\Pr\{n^{-1}\log N_{n,1}(\varepsilon, (\mathscr{F}_\delta')^2) > \tau\} = 0$ for all $\varepsilon, \tau, \delta > 0$. So, by Theorem 3.1 and (5.6),

$$\limsup_{n\to\infty} \Pr\big\{D_n^2(\delta) > 5\delta^2\big\} = 0. \tag{5.8}$$

By (5.5),

$$\begin{aligned} E\|n^{m/2}U_m^n(f)\|_{\mathscr{F}_\delta'} &\le E^*\bigg(\int_0^{5^{1/2}\delta}(\log N_{n,2}(\varepsilon, \mathscr{F}))^{m/2}\,d\varepsilon\bigg) \\ &\quad + E^*\bigg(\int_0^{\infty}(\log N_{n,2}(\varepsilon, \mathscr{F}))^{m/2}\,d\varepsilon\bigg) I_{D_n^2(\delta) > 5\delta^2}, \end{aligned}$$

and the equicontinuity condition for $n^{m/2}U_m^n$ follows from (5.8) and hypotheses (b) and (c). □

Theorem 5.6 has the following interesting corollary:

COROLLARY 5.7. *If $\mathscr{F}$ is an image admissible Suslin* VC *subgraph class of functions on S^m such that $P^mF^2 < \infty$, then all the projections $\pi_{k,m}\mathscr{F} = \{\pi_{k,m}f\colon f\in\mathscr{F}\}$, $k = 1,\dots,m$, satisfy the central limit theorem, that is,*

$$\big\{n^{k/2}U_k^n(\pi_{k,m}f)\colon f\in\mathscr{F}\big\} \to_{\mathscr{L}} \{K_{P,k}(\pi_{k,m}f)\colon f\in\mathscr{F}\}$$

in $l^\infty(\mathscr{F})$.

PROOF. If $\mathscr{F}$ is image admissible Suslin so is $\pi_{k,m}\mathscr{F}$ as mentioned in the Introduction. It also follows directly from the definitions that $(\mathscr{F}_\delta')^2$ and $[(\pi_{k,m}\mathscr{F})_\delta']^2$ are image admissible Suslin. So, the measurability hypotheses of Theorem 5.6 hold for the P-canonical classes $\pi_{k,m}\mathscr{F}$. Next we note that

$$\begin{aligned} e_{n,2}^2(\pi_{k,m}f, \pi_{k,m}g) &= U_k^n\Big(\big[\pi_{k,m}(f-g)\big]^2\Big) \\ &= U_k^n\Bigg(\bigg[\sum_{r=0}^{k}\frac{1}{r!}\sum_{(i_1,\dots,i_r)\in I_r^k}(-1)^{k-r} \\ &\qquad\qquad \times P^{m-r}(f-g)(x_{i_1},\dots,x_{i_r})\bigg]^2\Bigg) \\ &\le \sum_{r=0}^{k} c_{k,r}^2 U_n^r P^{m-r}(f-g)^2 \end{aligned}$$

for constants $c_{k,r}$ depending only on k and r. Hence,

$$N_{n,2}(\tau, \pi_{k,n}\mathscr{F}) \leq \prod_{r=0}^{k} N\left(\frac{\tau}{(k+1)^{1/2}c_{k,r}}, \mathscr{F}, \|\cdot\|_{L_2(U_n^r P^{m-r})}\right).$$

Since $\mathscr{F}$ is a VC subgraph, there are constants $c, v > 0$ such that for all probability measures Q on S^m with $QF^2 < \infty$ and all $\varepsilon > 0$,

$$N(\tau, \mathscr{F}, \|\cdot\|_{L_2(Q)}) \leq c\left[(QF^2)^{1/2}/\tau\right]^v$$

[Pollard (1984), Lemma 2.5 and Alexander (1987), Lemma 4.4]. It follows easily from the last two inequalities that the classes $\pi_{k,m}\mathscr{F}$ satisfy the hypotheses (b) and (c) of Theorem 5.6 (with k instead of m). Since they obviously satisfy (a), the result follows from Theorem 5.6. $\square$

This result contains Corollary 8 in Sherman (1991).

In Corollary 5.7 the hypothesis that $\mathscr{F}$ is a VC subgraph can obviously be replaced by: There is a function λ: $[0,\infty) \to [0,\infty)$ with $\int_0^\infty \lambda^{m/2}(\varepsilon)\,d\varepsilon < \infty$ such that for all probability measures Q on $(S, \mathscr{S})$ with $QF^2 < \infty$,

$$\log N_2\left(\varepsilon(QF^2)^{1/2}, \mathscr{F}, Q\right) \leq \lambda(\varepsilon), \qquad \varepsilon > 0.$$

6. Additional examples.

(a) *Discrete probabilities.* It is known [see, e.g., Dudley (1984), Theorem 6.3.1] that for discrete probabilities $P = \{p_i\}_{i=1}^\infty$ the empirical process indexed by the class of all subsets satisfies the CLT if and only if $\sum p_k^{1/2} < \infty$. We will see that this last condition implies the CLT for any uniformly bounded class, even in the degenerate case.

PROPOSITION 6.1. *Let P be a probability measure on $\mathbb{N}$ and let $p_k = P\{k\}$.*

(a) *If $\sum_k p_k^{1/2} < \infty$ and $\mathscr{F}$ is a uniformly bounded class of functions on $\mathbb{N}^m$, then*

$$n^{1/2}(U_m^n - P^m) \to_{\mathscr{L}} mG_P \circ S_m \quad \text{in } l^\infty(\mathscr{F}). \tag{6.1}$$

(b) *Conversely if* (6.1) *holds for the class of indicator functions of all the finite subsets of $\mathbb{N}^m$, then*

$$\sum_{k=1}^{\infty} p_k^{1/2} < \infty. \tag{6.2}$$

PROOF OF PROPOSITION 6.1(a). By Proposition 1.2, it suffices to show that the class of indicators of all the subsets of $\mathbb{N}^m$ satisfies the CLT. By Corollary 4.2, we only need to prove

$$n^{1/2}U_1^n(\pi_{1,m}S_mI_A) \to_{\mathscr{L}} mG_P(\pi_{1,m}S_mI_A) \quad \text{in } l^\infty(\mathscr{G}) \tag{6.3}$$

and

$$\|n^{1/2}U_k^n(\pi_{k,m}S_m I_A)\|_{\mathscr{F}} \to_{\Pr} 0, \qquad 2 \le k \le m. \tag{6.4}$$

Note that $n^{1/2}U_n^1(\pi_{1,m}S_m I_A) = \nu_n(P^{m-1}S_m I_A)$ and $\{P^{m-1}S_m I_A\colon A \subset \mathbb{N}^m\}$ is in the convex hull of $\{I_B\colon B \subset \mathbf{N}\}$. Hence (6.3) holds by the CLT for discrete measures in Dudley (1984) since the convex hull of a P-Donsker class is a P-Donsker class. As for (6.4), we just need to show

$$\sup_{A \subset \mathbb{N}^m} n^{1/2}|U_k^n(\pi_{k,m}I_A)| \to_{\Pr} 0.$$

Note that

$$\begin{aligned}
&E \sup_{A \subset N^m} \left|n^{1/2}U_k^n(\pi_{k,m}I_A)\right| \\
&\quad = E \sup_{A} \left| \sum_{(j_1,\dots,j_m)\in A} n^{1/2}U_k^n\big(\pi_{k,m}I_{\{(j_1,\dots,j_m)\}}\big)\right| \\
&\quad \le \sum_{j_1,\dots,j_m=1}^{\infty} E\left|n^{1/2}U_k^n\big(\pi_{k,m}I_{\{(j_1,\dots,j_m)\}}\big)\right| \\
&\quad \le \sum_{j_1,\dots,j_m=1}^{\infty} \left(E\left|n^{1/2}U_k^n\big(\pi_{k,m}I_{\{(j_1,\dots,j_m)\}}\big)\right|^2\right)^{1/2} \\
&\quad \le \sum_{j_1,\dots,j_m=1}^{\infty} \left|(k!(n-k)!/(n-1)!)P\{(j_1,\dots,j_m)\}\right|^{1/2} \\
&\quad = (k!(n-k)!/(n-1)!)^{1/2}\left(\sum_{j=1}^{\infty} p_j^{1/2}\right)^m \to_{n\to\infty} 0. \qquad \square
\end{aligned}$$

Proof of Proposition 6.1(b). If $A = B \times \mathbb{N}^{m-1}$ with $B \subset \mathbb{N}$, then $n^{1/2}(U_m^n - P^m)I_A = n^{1/2}(n^{-1}\sum_{i=1}^n I_{X_i \in B} - P(B))$. So (6.1) implies $\{\nu_n(I_B)\colon B \subset \mathbb{N}, B \text{ finite}\} \to_{\mathscr{L}} \{G_P(I_B)\colon B \subset \mathbb{N}, B \text{ finite}\}$. Hence $\sum_k p_k^{1/2} < \infty$ by Proposition 6.3.1 in Dudley (1985). □

Proposition 6.2. *Let P be a probability measure on $\mathbb{N}$ and let $p_k = P\{k\}$.*

(a) *If $\sum_k p_k^{1/2} < \infty$ and $\mathscr{F}$ is a uniformly bounded class of real functions on N^m, then*

$$n^{m/2}U_m^n \circ \pi_{m,m} \to_{\mathscr{L}} K_{P,m} \circ \pi_{m,m} \circ S_m \quad \text{uniformly in } l^\infty(\mathscr{F}).$$

(b) *Conversely, if $\{K_{P,m} \circ \pi_{m,m} \circ S_m I_A)\colon A \subset \mathbb{N}^m\}$ has bounded paths a.s., then* (6.2) *holds.*

Proof of Proposition 6.2(a). By the argument in Proposition 6.1 it is enough to consider the case $\{\pi_{m,m} \circ S_m(I_A)\colon A \subset \mathbb{N}^m\}$. We need to prove the usual equicontinuity condition

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left\{\sup_{e(A,B)\le\delta}\left|n^{m/2}U_m^n\pi_{m,m}(I_A-I_B)\right|>\varepsilon\right\}=0$$

for each $\varepsilon>0$, where $e^2(A,B)=P^m(A\,\triangle\,B)$. Note that this condition is implied by

$$\lim_{M\to\infty}\limsup_{n\to\infty} E\sup_{A\subset\mathbb{N}^m-[1,M]^m}\left|n^{m/2}U_m^n(\pi_{m,m}I_A)\right|=0. \tag{6.5}$$

We have

$$\begin{aligned}
E&\sup_{A\subset\mathbb{N}^m-[1,N]^m}\left|n^{m/2}U_m^n(\pi_{m,m}I_A)\right|\\
&\le mE\sup_{A\subset\mathbb{N}^m\cap[1,\infty)^{m-1}\times[N,\infty)}\left|n^{m/2}U_m^n(\pi_{m,m}I_A)\right|\\
&\le mE\sum_{i_1,\dots,i_{m-1}=1}^{\infty}\sum_{i_m=N}^{\infty}\left|n^{m/2}U_m\left(\pi_{m,m}I_{\{(i_1,\dots,i_m)\}}\right)\right|\\
&\le m\sum_{i_1,\dots,i_{m-1}=1}^{\infty}\sum_{i_m=N}^{\infty}\left(E\left|n^{m/2}U_m^n\left(\pi_{m,m}I_{\{(i_1,\dots,i_m)\}}\right)\right|^2\right)^{1/2}\\
&\le m\left(n^m(n-m)!/n!\right)^{1/2}\sum_{i_1,\dots,i_{m-1}=1}^{\infty}\sum_{i_m=N}^{\infty}\left(p_{i_1}\cdots p_{i_m}\right)^{1/2}.
\end{aligned}$$

Condition (6.5) follows from this estimate. □

PROOF OF PROPOSITION 6.2(b). As mentioned in the introduction,

$$n^{m/2}U_n^m h^{\phi}\to(\mathrm{Var}(\phi))^{m/2}(m!)^{-1/2}H_m\left(G_P(\phi)(\mathrm{Var}(\phi))^{-1/2}\right).$$

If $\phi_B=I_B-PB$, $B\subset\mathbb{N}$, we have

$$K_{P,m}(h^{\phi_B})=(\mathrm{Var}(I_B))^{m/2}(m!)^{-1/2}H_m\left(\sum_k p_k^{1/2}g_k(\delta_k-P)I_B(\mathrm{Var}(I_B))^{-1/2}\right)$$

since $G_P(I_B)=\sum_k p_k^{1/2}g_k(\delta_k-P)I_B$, with $\{g_k\}$ i.i.d. $N(0,1)$. Therefore, since H_m is a polynomial of degree m, the a.s. sample boundedness of $\{K_{P,m}(h^{\phi_B})$: $B\subset\mathbb{N}\}$ implies that of $\{\sum_k p_k^{1/2}g_k(\delta_k-P)I_B\colon B\subset\mathbb{N}\}$. Therefore $\sum_{k=1}^{\infty}p_k^{1/2}<\infty$. □

(b) *Sequences of functions.* Next we will give a result for U-processes that generalizes a well-known result for empirical processes.

PROPOSITION 6.3. *Let $\{h_k\}_{k=1}^{\infty}$ be a sequence of uniformly bounded functions from S^m into $\mathbb{R}$. Assume $Eh_k=0$ and*

$$\|h_k\|_\infty=o\left((1/\log k)^{1/2}\right). \tag{6.6}$$

Then $\{n^{1/2}U_m^n(h_k)\colon k\in\mathbb{N}\}\to_{\mathscr{L}}\{mG_P\circ P^{m-1}\circ S_m(h_k)\colon k\in\mathbb{N}\}$ in $l^\infty(\mathbb{N})$.

PROOF. Let $\mathscr{F} = \{h_k\}_{k=1}^{\infty}$. Since $\mathscr{F}$ is totally bounded with respect to the pseudometric $e_{P,m}$, we just need to check the equicontinuity condition

$$\lim_{N\to\infty} \limsup_{n\to\infty} P\Big\{ \sup_{k\ge N} n^{1/2}|U_m^n(h_k)| > \varepsilon \Big\} = 0.$$

By Proposition 2.3(c) and (6.6),

$$P\Big\{ \sup_{k\ge N} n^{1/2}|U_m^n(h_k)| > \varepsilon \Big\} \le \sum_{k=N}^{\infty} P\{n^{1/2}|U_m^n(h_k)| > \varepsilon\}$$
$$\le \sum_{k=N}^{\infty} c\exp\big(-c'\varepsilon^2\|h_k\|^{-2}\big),$$

which converges to zero. □

In the canonical case an analogous result can be obtained using (d) of Proposition 2.3:

PROPOSITION 6.4. *Let $\{h_k\}_{k=1}^{\infty}$ be a sequence of canonical functions from S^m into $\mathbb{R}$. Assume $Eh_k = 0$ and $\|h_k\|_\infty = o((1/\log k)^{m/2})$. Then*

$$\{n^{m/2}U_m^n(h_k) \colon k \in \mathbb{N}\} \to_{\mathscr{L}} \{K_{P,m} \circ S_m(h_k) \colon k \in \mathbb{N}\} \quad in\ l^\infty(\mathbb{N}).$$

REMARK 6.5. As in the empirical process case these two propositions are best possible. Consider a double sequence $\{\varepsilon_{i,j}\}_{i,j=1}^{\infty}$ of i.i.d. r.v.'s with $P\{\varepsilon_{i,j} = 1\} = P\{\varepsilon_{i,j} = -1\} = 1/2$. Take S to be the unit ball of l_∞, $X_i = (\varepsilon_{i_1}, \varepsilon_{i_2}, \dots)$ for each i. Define $f_k, h_k \colon S^m \to \mathbb{R}$ by $f_k(x_1, \dots, x_m) = x_1^{(k)}(\log k)^{-1/2}$ if $k \ge 2$, $f_1(x_1, \dots, x_m) = 0$, $h_k(x_1, \dots, x_m) = x_1^{(k)} \cdots x_m^{(k)}(\log k)^{-m/2}$ if $k \ge 2$ and $h_1(x_1, \dots, x_m) = 0$, where $x_j^{(k)}$ is the kth coordinate of x_j. Then $\{f_k\}_{k=1}^{\infty}$ is a sequence of uniformly bounded functions from S^m into $\mathbf{R}$ with $Ef_k = 0$, $\|f_k\|_\infty = (\log k)^{-1/2}$, $k \ge 2$, and we claim that $\{f_k\}_{k=1}^{\infty}$ does not satisfy the CLT. Since $n^{1/2}U_m^n(f_k) = (\log k)^{-1/2}n^{-1/2}\sum_{i=1}^n \varepsilon_{ik}$ for $k \ge 2$, with $\{g_k\}_{k=2}^{\infty}$ i.i.d. r.v.'s with law $N(0,1)$ and $g_1 = 0$, the finite-dimensional distributions of the process $\{n^{1/2}U_m^n(f_k)\}_{k=1}^{\infty}$ converge to $\{(\log k \vee 2)^{-1/2}g_k\}_{k=1}^{\infty} =: \{G(k)\}_{k=1}^{\infty}$. But the paths of this process are not continuous at $k = 1$: $e(k,1) \to 0$ as $k \to \infty$, however, since $|G(k) - G(1)| = (\log k)^{-1/2}|g_k|$ and $\lim_{t\to\infty} t^{-2}\log P\{|g| \ge t\} = -1/2$, the Borel–Cantelli lemma gives $\limsup(\log k)^{-1/2}|g_k| = 2^{1/2}$ a.s.

The situation for $\{h_k\}_{k=1}^{\infty}$ is similar. Let H_k be the kth Hermite polynomial. By the CLT for degenerate U-statistics in Rubin and Vitale (1980), $n^{m/2}U_n^m(h_k) \to_w H_m(g_k)(\log k)^{-m/2}(m!)^{-1/2}$ for $k \ge 2$. Since $H_m(g_k)$ is a polynomial in m variables with leading coefficient equal to 1,

$$\lim_{t\to\infty} t^{-2/m}\log P\{|H_m(g_k)| \ge t\} = -1/2.$$

Hence $\limsup(\log k)^{-m/2}H_m(g_k)(m!)^{-1/2} = 2^{m/2}(m!)^{-1/2}$ a.s.

(c) *The simplicial depth process.* Given a probability measure P on $\mathbb{R}^k$, the simplicial depth process on $\mathbb{R}^k$ is defined by Liu (1990) as $D_P(x) = P^{k+1}\{x \in$

$S(x_1, \dots, x_{k+1})\}$, where $S(x_1, \dots, x_{k+1})$ is the closed simplex determined by the points $x_1, \dots, x_{k+1} \in \mathbb{R}^k$. The empirical simplicial depth for a sample $X_1, \dots, X_n$ from P is

$$D_n(x) = U^n_{k+1}(C_x) \quad \text{where } C_x = I_{\{(x_1, \dots, x_{k+1}):\, x \in S(x_1, \dots, x_{k+1})\}}.$$

So, for each x, C_x is a function defined on $(\mathbb{R}^k)^{k+1}$ or more specifically, on simplices of $\mathbb{R}^k$. We first observe that $\mathscr{F} = \{C_x : x \in \mathbb{R}^k\}$ is the collection of indicators of a measurable Vapnik–Červonenkis class of sets. To this end we recall a well-known lemma [Vapnik and Červonenkis (1971)].

LEMMA 6.6. *If $P_k(n)$ denotes the number of sets of the form $\tilde{A}_1 \cap \cdots \cap \tilde{A}_n$, where $\tilde{A}_i$ is either A_i or A_i^c and $A_1, \dots, A_n$ are n halfspaces of $\mathbb{R}^k$, then*

$$P_k(n) \le \sum_{j=0}^{k \wedge n} \binom{n}{j}. \tag{6.7}$$

COROLLARY 6.7. *$\mathscr{F}$ is the collection of indicator functions of a measurable Vapnik–Červonenkis class of sets.*

PROOF. Obviously the map $(x, x_1, \dots, x_{k+1}) \to I_{C_x}(x_1, \dots, x_{k+1})$ is jointly Borel measurable, hence $\mathscr{F}$ is image admissible Suslin and therefore, measurable. Let $S_1, \dots, S_n$ be n simplices in $\mathbb{R}^k$. Each is defined by $k + 1$ halfspaces, therefore they divide $\mathbb{R}^k$ into at most $P_k((k + 1)n)$ subsets. But $\Delta^{\mathscr{F}}(S_1, \dots, S_n) = \#[\{S_1, \dots, S_n\} \cap C : I_C \in \mathscr{F}]$ is precisely the number of subsets that $S_1, \dots, S_n$ determine on $\mathbb{R}^k$, that is, $\Delta^{\mathscr{F}}(S_1, \dots, S_n) \le P_k((k + 1)n)$. By Lemma 6.6, $P_k((k + 1)n)$ is a polynomial in n of degree k. So, $\mathscr{F}$ is VC [see, e.g., Dudley (1986) for the definition of a VC class]. □

Now we can state the law of large numbers and the central limit theorem for the simplicial depth process. Dümbgen (1990) proved the LLN and the CLT for simplicial depth processes under slightly less generality and Liu (1990) has an earlier version of the LLN under additional hypotheses.

COROLLARY 6.8. *For all P in $\mathbb{R}^k$,*

$$\sup_x |D_n(x) - D_P(x)| \to 0 \quad a.s.,$$

and

$$\mathscr{L}\big(n^{1/2}(D_n(x) - D_P(x))\big) \to_w \mathscr{L}\big(kG_P \circ P^{k-1}\big) \quad in\ l^\infty(\mathbb{R}^k).$$

PROOF. This follows from Corollaries 3.3 and 4.9, together with Corollary 6.7. □

The law of large numbers in Corollary 6.8 yields consistency of the empirical simplicial median under minimal conditions [see Liu (1990), Theorem 5, for a less general version].

THEOREM 6.9. *Let P be a probability measure on $\mathbb{R}^k$ satisfying*:

(a) *$D(\cdot)$ is uniquely maximized at μ.*

(b) *μ_n is a sequence of random variables with $D_n(\mu_n) = \sup_x D_n(x)$.*

Then

$$\mu_n \to \mu \quad a.s.$$

PROOF. First we see that $D(\cdot)$ is an upper semicontinuous function: If $y_n \to y$, then $\limsup D(y_n) \le P^{k+1}[\limsup\{y_n \in S(x_1, \dots, x_{k+1})\}] \le P^{k+1}\{y \in S(x_1, \dots, x_{k+1})\}$. Since $D(\cdot)$ is an upper semicontinuous function and $\lim_{|x| \to \infty} D(x) = 0$ [Theorem 1, Liu (1990)] it follows that $\delta =: D(\mu) - \sup_{|x-\mu| \ge \varepsilon} D(x) > 0$. Hence,

$$\begin{aligned} P\Big\{\sup_{n \ge l} |\mu_n - \mu| > \varepsilon\Big\} &\le P\Big\{\sup_{n \ge l}(D(\mu) - D(\mu_n)) \ge \delta\Big\} \\ &\le P\Big\{\sup_{n \ge l}(D(\mu) - D_n(\mu)) + \sup_{n \ge l}(D_n(\mu_n) - D(\mu_n)) \ge \delta\Big\} \\ &\le 2P\Big\{\sup_{n \ge l} \sup_x |D_n(x) - D(x)| \ge \delta/2\Big\} \to 0 \end{aligned}$$

as $l \to \infty$ by Corollary 6.8. □

If the simplices $S(x_1, \dots, x_{k+1})$ in the definition of $\mathscr{F}$ are taken to be *open*, then $\mathscr{F}$ is still a measurable VC class and Corollary 6.8 holds. Theorem 6.9 also holds for the corresponding definitions of μ and μ_n if it is further assumed that P gives zero mass to hyperplanes: Under this hypothesis the function D is the same as for closed simplices, hence, upper semicontinuous.

REFERENCES

ALEXANDER, K. S. (1987). The central limit theorem for empirical processes on Vapnik–Červonenkis classes. *Ann. Probab.* **15** 178–203.

ANDERSEN, N. T. and DOBRIĆ, V. (1987). The central limit theorem for stochastic processes. *Ann. Probab.* **15** 164–177.

ANDERSEN, N. T., GINÉ, E., OSSIANDER, M. and ZINN, J. (1988). The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. *Probab. Theory Related Fields* **77** 271–305.

ARCONES, M. A. (1991). Limits of canonical U-processes. Preprint.

ARCONES, M. A. and GINÉ, E. (1991). On decoupling, series expansions and tail behavior of chaos processes. *J. Theoret. Probab.* To appear.

BONAMI, A. (1970). Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier* **20** 335–402.

BOREL, C. (1979). On the integrability of Banach space valued Walsh polynomials. *Séminaire de Probabilités XIII. Lecture Notes in Math.* **721** 1–3. Springer, New York.

BORISOV, I. S. (1990). Exponential inequalities for the distribution of von Mises and U-statistics. Preprint.

BRETAGNOLLE, J. (1983). Lois limites du bootstrap de certaines fonctionelles. *Ann. Inst. H. Poincaré Sect. B.* **3** 256–261.

BRONSTEIN, E. M. (1976). ε-entropy of convex sets and functions. *Siberian Math. J.* **17** 393–398.

CARL, A. and PAJOR, A. (1988). Gelfand numbers of operators in a Hilbert space. *Invent. Math.* **94** 479–504.

DE LA PEÑA, V. H. (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.* **20** 1877–1892.

DUDLEY, R. M. (1967). The size of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1** 290–330.

DUDLEY, R. M. (1984). A course on empirical processes. *École d'Eté de Probabilités de Saint-Flour XII. Lecture Notes in Math.* **1097** 1–142. Springer, New York.

DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical functions. *Probability in Banach Spaces V. Lecture Notes in Math.* **1153** 141–178. Springer, New York.

DUDLEY, R. M. (1989). *Real Analysis and Probability.* Wadsworth and Brooks Cole, Pacific Grove, CA.

DUDLEY, R. M., GINÉ, E. and ZINN, J. (1991). Uniform and universal Glivenko-Cantelli classes of functions. *J. Theoret. Probab.* **4** 485–510.

DÜMBGEN, L. (1990). Limit theorems for the empirical simplicial depth. *Statist. Probab. Lett.* **14** 119–128.

DYNKIN, E. B. and MANDELBAUM, A. (1983). Symmetric statistics, Poisson point processes and multiple Wiener integrals. *Ann. Statist.* **11** 739–745.

FERNIQUE, X. (1983). Régularité des fonctions aléatoires non-Gausiennes. *École d'Eté de Probabilités de Saint-Flour XI. Lecture Notes in Math.* **976** 1–74. Springer, New York.

GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–989.

GINÉ, E. and ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Lecture Notes in Math.* **1221** 50–113. Springer, New York.

GINÉ, E. and ZINN, J. (1992a). Marcienkiewicz type laws of large numbers and convergence of moments for U-statistics. In *Probability in Banach Spaces* **8** 273–291. Birkhäuser, Boston.

GINÉ, E. and ZINN, J. (1992b). On Hoffmann-Jørgensen's inequality for U-processes. In *Probability in Banach Spaces* **8** 80–91. Birkhäuser, Boston.

GREGORY, G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5** 110–123.

HELMERS, J., JANSSEN, P. and SERFLING, R. (1988). Glivenko–Cantelli properties of some generalized empirical df's and strong convergence of generalized L-statistics. *Probab. Theory Related Fields* **79** 75–93.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HOFFMANN-JØRGENSEN, J. (1984). Stochastic processes on Polish spaces. Aarhus Univ., Mat. Inst. Various Publications Series, No. 39. Aarhus, Denmark. Unpublished manuscript.

KAHANE, J. P. (1968). *Some Random Series of Functions.* D. C. Heath, Lexington, MA.

KWAPIEŃ, S. (1987). Decoupling for polynomial chaos. *Ann. Probab.* **15** 1062–1071.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer, New York.

LIU, R. V. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414.

NELSON, E. (1973). The free Markov field. *J. Funct. Anal.* **12** 211–227.

NOLAN, D. and POLLARD, D. (1987). U-Processes: rates of convergence. *Ann. Statist.* **15** 780–799.

NOLAN, D. and POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16** 1291–1298.

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897–919.

PISIER, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Banach Spaces, Harmonic Analysis, and Probability Theory: Proceedings of the Special Year in Analysis. Lecture Notes in Math.* **995** 123–154. Springer, New York.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

RUBIN, M. and VITALE, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** 165–170.

SCHNEEIMEIER, W. (1989). Weak convergence and Glivenko-Cantelli results for empirical processes of U-statistics structure. *Stochastic Process. Appl.* **33** 325–334.

SEN, P. K. (1974). On L_p convergence of U-statistics. *Ann. Inst. Statist. Math.* **26** 55–60.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SHERMAN, R. P. (1991). Maximal inequalities for degenerate U-processes with applications to optimization estimators. Preprint, Bell Communications Research, Murray Hill, NJ.

STRASSEN, V. and DUDLEY, R. M. (1969). The central limit theorem and ε-entropy. *Probability and Information Theory. Lecture Notes in Math.* **89** 224–231.

TALAGRAND, M. (1987). The Glivenko–Cantelli problem. *Ann. Probab.* **15** 837–870.

VAPNIK, V. N. and ČERVONENKIS. A. JA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 164–280.

VAPNIK, V. N. and ČERVONENKIS, A. JA. (1981). Necessary and sufficient conditions for the convergence of means to their expectation. *Theory Probab. Appl.* **26** 532–553.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UTAH 84112

DEPARTMENT OF MATHEMATICS, U-9
UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT 06269