Proc. Probab. Banach Spaces IX (Sandbjerg 1993), *Progr. Probab.* Vol. 35, Birkhäuser Boston, 1994, (104-127)
Preprint Ser. No. 15, 1993, *Math. Inst. Aarhus*

Uniform Ergodic Theorems for Dynamical Systems Under VC Entropy Conditions

GORAN PESKIR *and* JOSEPH E. YUKICH*

The classic limit theorems of Vapnik and Chervonenkis [27,28] show that if a function class $\mathcal{F}$ satisfies a random entropy condition, then the strong law of large numbers holds uniformly over $\mathcal{F}$. In this paper we show that an analogous weighted entropy condition implies that Birkhoff's pointwise ergodic theorem holds uniformly over $\mathcal{F}$. In this way we obtain a variety of uniform ergodic theorems for measure-preserving semi-flows. The method of proof relies upon a blocking technique, a decoupling inequality due to Eberlein, and a standard subgaussian inequality for Rademacher averages. The results extend and generalize to provide uniform ergodic theorems for operators.

1. Introduction. Let T be a measure-preserving transformation of the probability space $(\Omega, \mathcal{A}, \mu)$. The classic pointwise ergodic theorem of Birkhoff states that if $f \in L^1(\mu)$, then the sequence of time averages $n^{-1} \sum_{i=0}^{n-1} f(T^i(\omega))$ converges for μ-almost all $\omega \in \Omega$. If T is ergodic, then the limiting value is the space average $\int_\Omega f \; d\mu$, that is we have:

$$(1.1) \qquad \lim_{n\to\infty} \Big| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(\omega)) - \int_\Omega f \, d\mu \Big| = 0 \quad \text{for } \mu\text{-a.s. } \omega \in \Omega .$$

The present paper is motivated by the following question: When does the ergodic theorem (1.1), as well as its generalization to measurable measure-preserving semi-flows $\{T_t\}_{t\ge 0}$, hold uniformly over function classes $\mathcal{F} \subset L^1(\mu)$ with μ arbitrary? In other words, when does a uniform ergodic theorem of the type:

$$(1.2) \qquad \lim_{n\to\infty} \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(\omega)) - \int_\Omega f \, d\mu \Big| = 0 \quad \text{for } \mu\text{-a.s. } \omega \in \Omega$$

hold whenever μ is a probability measure? More generally, if T and P are operators in $L^1(\mu)$ under what conditions on the function class $\mathcal{F}$ of maps from $\mathbf{R}$ into $\mathbf{R}$ and a given function $g \in L^1(\mu)$ does it follow that:

$$(1.3) \qquad \lim_{n\to\infty} \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} \Big(f(T^i(g)(\omega)) - P\big(f(T^i(g))\big)(\omega)\Big) \Big| = 0 \quad \text{for } \mu\text{-a.s. } \omega \in \Omega \ ?$$

If P is defined by $P(h) = \int h \; d\mu$ for $h \in L^1(\mu)$, then (1.3) reduces to finding sufficient conditions insuring the uniform ergodic theorem of the form:

*Research partially supported by an NSF grant.
AMS 1980 *subject classifications*. Primary 28D05, 28D10, 47A35, 60G10. Secondary 60B12, 60F15, 60F17, 60F25.
Key words and phrases: Uniform ergodic theorems, Vapnik-Chervonenkis entropy numbers, absolute regularity (β-mixing), measurable dynamical systems, decoupling, moving averages.

(1.4) $$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} \Big(f\big(T^i(g)(\omega)\big) - \int_\Omega f\big(T^i(g)(\omega)\big)\,\mu(d\omega)\Big) \Big| = 0 \ \text{ for } \mu\text{-a.s. } \omega \in \Omega \ .$$

Although the results (1.3) and (1.4) may be approached from many directions, we will find it convenient to approach them through the uniform ergodic theorem (1.2). Taking this approach, identify the sequence of measure-preserving transformations $\{ T^i \}_{i=0}^{n-1}$ with a stationary sequence of random variables $\{ X_i \}_{i=1}^{n}$, where the law of X_1 equals μ . Letting $\mu_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ be the usual empirical measures for μ , the uniform ergodic theorem (1.2) in its simplified form may be reduced to finding conditions on $\mathcal{F}$ such that the following uniform strong law of large numbers (SLLN) holds:

(1.5) $$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \Big| \int_\Omega f \ d(\mu_n - \mu) \Big| = 0 \quad \mu\text{-a.s. } \omega \in \Omega$$

whenever μ is a probability measure.

Since the seminal work of Vapnik and Chervonenkis [27,28], it is now well-known that if the random variables $\{ X_i \}_{i\geq 1}$ are independent and identically distributed (i.i.d.) with the fixed but arbitrary distribution μ , and if $\mathcal{F}$ is a VC subgraph class of functions, see Dudley [6], then the uniform Glivenko-Cantelli theorem (1.5) holds. By considering what amounts to random entropy numbers they proved the following remarkable SLLN. By entropy number $N_n(\varepsilon, \mathcal{F})$ we mean the least number of open balls in the sup-metric of radius $\varepsilon > 0$ which form a covering of the set of all vectors in $\mathbf{R}^n$ of the form $\big(f(X_1), \ldots, f(X_n)\big)$ where f ranges over $\mathcal{F}$, and where $n \geq 1$ is given and fixed. Here and henceforth we shall assume that $N_n(\varepsilon, \mathcal{F})$ is measurable.

Theorem 1.1 (**Vapnik and Chervonenkis** [27,28]) *Let $\mathcal{F} \subset L^1(\mu)$ be a uniformly bounded class of functions. Then $\mathcal{F}$ satisfies the uniform strong law of large numbers (1.5) for all probability measures μ if and only if for all $\varepsilon > 0$:*

(1.6) $$\lim_{n\to\infty} \frac{E \log N_n(\varepsilon, \mathcal{F})}{n} = 0 \ .$$

Actually, Vapnik and Chervonenkis proved the equivalence of (1.6) and the convergence in probability version of (1.5). Steele [25] showed that this convergence is a.s.; see also Kuelbs and Zinn [14] and Pollard [22]. This result, together with the general theory of VC classes, has stimulated substantial areas of empirical processes theory and has had surprising consequences in fields ranging from Banach space theory to statistics, see e.g. Dudley [5], Gaenssler [8], Giné-Zinn [9], and Pollard [23]. It is now well-known, for example, that central limit theorems (CLT's) for the function-indexed empirical process $\int f \ d(\mu_n - \mu)$, $f \in \mathcal{F}$, hold for all probability measures μ whenever $\mathcal{F}$ is a VC subgraph class of functions.

The purpose of this paper is to show that VC classes play a similarly important role in the context of uniform ergodic theorems. *In fact the main results of this paper (Theorem 3.1, Theorem 3.2 and Theorem 4.1) show that a weighted entropy condition, representing a natural generalization of (1.6), implies that the uniform ergodic theorems (1.2) and (1.4) hold for all μ .*

We do not wish to review the history of Theorem 1.1 and its extensions and generalizations, but will point out the fundamental papers of Giné and Zinn [9] and

Talagrand [26]. Giné and Zinn [9] showed how to effectively use randomization as a tool for proving the uniform SLLN as well as the CLT in the i.i.d. case. Randomizing in (1.5) with a Rademacher sequence $\{\varepsilon_j\}_{j\geq 1}$, one obtains the simple but fundamental estimate:

$$E \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} \big(f(X_i) - Ef(X_i)\big) \Big| \leq 2\, E \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \cdot f(X_i) \Big|$$

for all $n \geq 1$. This estimate, together with the fact that $\sum_{i=1}^{n} \varepsilon_i \cdot f(X_i)$, $f \in \mathcal{F}$, represents a subgaussian process, form the key contribution of randomization.

Unfortunately, such an estimate fails to hold for stationary ergodic sequences $\{X_i\}_{i\geq 1}$, see Peškir and Weber [18]. However, when the sequence $\{X_i\}_{i\geq 1}$ has an additional weak dependence structure involving a form of mixing, then this difficulty may be circumvented via decoupling inequalities and blocking techniques. Blocking, which goes back to Bernstein, has proved useful in many instances, notably in the work of Philipp [20], Eberlein [7], Yukich [33], Massart [15] and more recently the relevant paper of Yu [31].

In the present paper we use the blocking technique for sequences of absolutely regular (β-mixing) random variables, together with the subgaussian inequality to show that a weighted VC entropy condition implies the uniform ergodic theorem (1.2). Our sufficient conditions insuring (1.2) are simple and moreover, the proof is straightforward. The recent interesting and more technically complicated work of Yu [31], and Arcones and Yu [2], provides rates of convergence for the uniform SLLN (1.5) as well as CLT's. The approach of the present paper makes no attempt in this direction but could be modified to treat these questions. To conclude the introduction we clarify that the measurability of functions under consideration is implicitly assumed wherever needed. Recall the well-known fact that this approach might be supported in quite a general setting by using the theory of analytic spaces. For more details in this direction we refer the reader to Yu's paper [31].

2. The VC law of large numbers in the stationary case.

The aim of this section is to generalize and extend the VC law of large numbers to stationary random variables. In some sense, the results and methods of this section could be interpreted as both a refinement and simplification of those in Yu's recent paper [31]. Throughout, let $\{X_i\}_{i\geq 1}$ be a *stationary* sequence of random variables defined on the probability space $(\Omega, \mathcal{A}, P)$, with values in a measurable space $(S, \mathcal{S})$ and common distribution law π , and with distribution law μ in $(S^{\mathbf{N}}, \mathcal{S}^{\mathbf{N}})$. More precisely, this means that:

$$(X_{n_1}, \ldots, X_{n_k}) \sim (X_{n_1+\tau}, \ldots, X_{n_k+\tau})$$

for all $1 \leq n_1 < \ldots < n_k$ and all $\tau \geq 1$. We recall that the stationary sequence $\{X_i\}_{i\geq 1}$ is called *ergodic*, if the unilateral shift $\Theta : S^{\mathbf{N}} \to S^{\mathbf{N}}$ defined by:

$$\theta(s_1, s_2, s_3, \ldots) = (s_2, s_3, \ldots)$$

is ergodic with respect to μ . For every $l \geq 1$ introduce the σ-algebras:

$$\sigma_1^l = \sigma(X_1, \ldots, X_l) \;\text{ and }\; \sigma_l^\infty = \sigma(X_{l+1}, X_{l+2} \ldots) \; .$$

Define the β-mixing coefficients of the sequence $\{X_i\}_{i\geq 1}$ by:

(2.1) $$\beta_k = \beta_k\big(\{ X_i \}_{i\geq 1}\big) = \sup_{l\geq 1} \int \sup_{A\in\sigma_{k+l}^{\infty}} |P(A \,|\, \sigma_1^l) - P(A)| \, dP$$

for all $k \geq 1$. Equivalently, the β-mixing coefficients may be defined as follows, see [3]:

(2.1') $$\beta_k = \sup \Big\{ \textstyle\sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)\cdot P(B_j)| : (A_i)_{i=1}^{I} \text{ is any finite partition in } \sigma_1^l \text{, and } (B_j)_{j=1}^{J} \text{ is any finite partition in } \sigma_{k+l}^{\infty} \text{ for } I, J, l \geq 1 \Big\}$$

for all $k \geq 1$.

The sequence $\{ X_i \}_{i\geq 1}$ is called *absolutely regular* (β-mixing), if $\beta_k \to 0$ as $k \to \infty$. The concept of absolute regularity was first studied by Volkonskii and Rozanov [29,30] who attribute it to Kolmogorov.

It is well-known that if the sequence $\{ X_i \}_{i\geq 1}$ is absolutely regular, then it is strongly mixing, and therefore ergodic, see [19] (p.57). Thus, by Birkhoff's theorem if $X_1 \in L^1(P)$, then the following strong law of large numbers holds:

$$\frac{1}{n} \sum_{i=1}^{n} \big(X_i - EX_i\big) \to 0 \quad P\text{-}a.s.$$

as $n \to \infty$. It should be noticed that since the sequence $\{ X_i \}_{i\geq 1}$ is assumed to be stationary, then all random variables X_i are identically distributed for $i \geq 1$, and therefore we have $EX_i = EX_1$ for all $i \geq 1$. By the same argument it follows that, if $f \in L^1(\pi)$ then we have:

$$\frac{1}{n} \sum_{i=1}^{n} \big(f(X_i) - Ef(X_i)\big) \to 0 \quad P\text{-}a.s.$$

as $n \to \infty$, with $Ef(X_i) = Ef(X_1)$ for all $i \geq 1$.

It is our wish to extend this SLLN and obtain a uniform SLLN over a class $\mathcal{F}$ of real valued functions on S . Although Peškir and Weber [17] have recently characterized the uniform SLLN in the general setting of stationarity, our goal here is to provide an analog of the classic VC law of large numbers [27,28] in the setting of absolute regularity. This approach involves conditions on the entropy number for $\mathcal{F}$.

By *entropy number* we mean $N_n^X(\varepsilon, \mathcal{F})$, which denotes the smallest number of open balls in the sup-metric of radius $\varepsilon > 0$ which form a covering of the set of all vectors in $\mathbf{R}^n$ of the form $\big(f(X_1), \ldots, f(X_n)\big)$ where f ranges over $\mathcal{F}$, and where $n \geq 1$ is given and fixed.

The main result of this section may now be stated as follows. *Later, in section 4, we will see that a P-probability version of Theorem 2.1 holds for sequences of random variables which are not stationary; in this way we deduce the uniform ergodic theorem (1.4).*

Theorem 2.1. *Let $\{X_i\}_{i\geq 1}$ be an absolutely regular sequence of random variables satisfying the condition:*

(2.2) $$w_n^{-1} \beta_{w_n} = o(n^{-1})$$

for some sequence $w_n = o(n)$. If $\mathcal{F}$ is a uniformly bounded class of functions satisfying:

$$(2.3) \qquad \lim_{n\to\infty} \; w_n \; \frac{E \log N_n^X(\varepsilon, \mathcal{F})}{n} \; = \; 0$$

for all $\varepsilon > 0$ *, then* $\mathcal{F}$ *satisfies the uniform strong law of large numbers as follows:*

$$(2.4) \qquad \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} \big(f(X_i) - Ef(X_i) \big) \Big| \longrightarrow 0 \;\; P\text{-}a.s.$$

as $n \to \infty$.

Remarks.

1. It is easily verified that condition (2.2) with some sequence $w_n = o(n)$ is equivalent to the following condition:

$$(2.2') \qquad \tfrac{n}{2} - w_n < b_n w_n \leq \tfrac{n}{2} \quad \& \quad b_n \beta_{w_n} \to 0$$

where $b_n = [n/2w_n]$ for $n \geq 1$. It turns out that (2.2') is precisely a version of condition (2.2) which will be used in the proof below. When X_i , $i \geq 1$ are i.i.d., we may take $w_n = 1$ for all $n \geq 1$, and since in this case $\beta_k = 0$ for all $k \geq 1$, we see that (2.2) is satisfied. Moreover, the weighted entropy condition (2.3) in this case reduces to the classical VC entropy condition. In this way we recover the sufficiency part of the VC Theorem 1.1. Finally, since the sequence $\{ \beta_k \}_{k\geq 1}$ is decreasing, it is no restriction to assume in Theorem 2.1 that $w_n \to \infty$, for otherwise the β-mixing coefficients β_k , $k \geq 1$ are eventually identically zero, so we are in the setting of the classic VC theorem. To see this, assume that w_n does not tend to infinity as $n \to \infty$. Then there exist a subsequence $\{ w_{n_k} \}_{k\geq 1}$ of $\{ w_n \}_{n\geq 1}$ and $N \geq 1$ such that $w_{n_k} \leq N$ for all $k \geq 1$. Suppose that (2.2) holds. Then we have:

$$n_k \frac{\beta_N}{N} \leq n_k \cdot \frac{\beta_{n_k}}{w_{n_k}} \to 0$$

as $k \to \infty$. Therefore $\beta_n = 0$ for all $n \geq N$, and the claim follows.

2. When $\mathcal{F}$ is a VC class of functions, the conclusion (2.4) of Theorem 2.1 holds whenever there exists a sequence $w_n = o(n)$, such that:

$$(2.5) \qquad n \cdot \frac{\beta_{w_n}}{w_n} \to 0 \quad \& \quad w_n \cdot \frac{\log n}{n} \to 0$$

as $n \to \infty$. For example, consider the case when *the mixing rate* r_β is strictly positive, where we recall that $r_\beta = \sup \{ r \geq 0 \mid \{n^r \beta_n\}_{n\geq 1} \text{ is bounded } \}$. Then $n^r \beta_n \to 0$ for some $r > 0$. Put $w_n = n^{1/(1+r)}$ for $n \geq 1$. Then we clearly have:

$$n \cdot \frac{\beta_{w_n}}{w_n} = w_n^r \cdot \beta_{w_n} \to 0 \quad \& \quad w_n \cdot \frac{\log n}{n} \to 0$$

as $n \to \infty$. Thus (2.5) is satisfied. Therefore if $\mathcal{F}$ is a bounded VC class of functions then the uniform SLLN holds whenever the mixing rate is positive. This improves upon the recent work of Arcones and Yu (Corollary 3 of [2]), who show that the CLT and thus the uniform SLLN hold whenever the mixing rate is greater than 1.

Before proving Theorem 2.1 we establish some preliminary results. The proof is centered around the blocking technique as follows. Setting $b_n = [n/2w_n]$ for $n \geq 1$, divide the sequence $(X_1, \dots, X_n)$ into $2b_n$ blocks of length w_n , leaving a remainder block of length $n - 2b_n w_n$. Define blocks:

$$B_j = \{ \ i \ | \ 2(j-1)\, w_n + 1 \leq i \leq (2j-1)\, w_n \ \}$$

$$\hat{B}_j = \{ \ i \ | \ (2j-1)\, w_n + 1 \leq i \leq 2j\, w_n \ \}$$

$$R = \{ \ i \ | \ 2b_n w_n + 1 \leq i \leq n \ \}$$

for all $1 \leq j \leq b_n$. Using the above blocks, define a sequence $\{ Y_i \}_{i \geq 1}$ of random variables on a probability space $(\Lambda, \mathcal{B}, Q)$ with values in the measurable space $(S, \mathcal{S})$ and coupled to the sequence $\{ X_i \}_{i \geq 1}$ by the relation:

$$\mathcal{L}(Y_1, \dots, Y_{b_n w_n}) = \bigotimes_{1}^{b_n} \mathcal{L}(X_1, \dots, X_{w_n})$$

for all $n \geq 1$. The next lemma, first noticed by Eberlein [7], compares the original sequence $\{X_i\}_{i \geq 1}$ with the coupled block sequence $\{Y_i\}_{i \geq 1}$. This lemma, which may be interpreted as a decoupling inequality with an error term, plays a central role in the sequel.

Lemma 2.2. *The following estimate is valid:*

$$\Big| P\{ (X_1, \dots, X_{w_n}, X_{2w_n+1}, \dots, X_{3w_n}, \ \dots \ , X_{(2b_n-2)w_n+1}, \dots, X_{(2b_n-1)w_n}) \in B \}$$
$$- Q\{ (Y_1, \dots, Y_{w_n}, Y_{2w_n+1}, \dots, Y_{3w_n}, \ \dots \ , Y_{(2b_n-2)w_n+1}, \dots, Y_{(2b_n-1)w_n}) \in B \} \Big|$$
$$\leq \ (b_n - 1) \cdot \beta_{w_n}$$

for all measurable sets $B \in \bigotimes_1^{b_n} \mathcal{S}$ *, and all* $n \geq 1$ *.*

Proof. It follows from (2.1'), the monotone class lemma, and induction, see [7]. □

By Lemma 2.2 it follows that for any bounded measurable function $g : S^{b_n w_n} \rightarrow \mathbf{R}$ we have the decoupling estimate:

(2.6)
$$\Big| Eg\,(X_1, \dots, X_{w_n}, X_{2w_n+1}, \dots, X_{3w_n}, \ \dots \ , X_{(2b_n-2)\,w_n+1}, \dots, X_{(2b_n-1)w_n})$$
$$- Eg\,(Y_1, \dots, Y_{w_n}, Y_{2w_n+1}, \dots, Y_{3w_n}, \ \dots \ , Y_{(2b_n-2)\,w_n+1}, \dots, Y_{(2b_n-1)w_n}) \Big|$$
$$\leq \ (b_n - 1) \cdot \beta_{w_n} \cdot \| \, g \, \|_\infty$$

for all $n \geq 1$. The idea of using Eberlein's Lemma 2.2 together with the above blocking technique is apparently due to Yu [31].

Lemma 2.3. *Let* $\{ Z_i \mid i \geq 1 \}$ *be a stationary ergodic sequence of random variables defined on the probability space* $(\Omega, \mathcal{A}, P)$ *, with values in a measurable space* $(S, \mathcal{S})$ *, and with common distribution law* π *. Let* $\mathcal{F}$ *be a class of real valued functions on* S *such that the envelope* $F_\mathcal{F}(s) = sup_{f \in \mathcal{F}} | \, f(s) \, |$ *of* $\mathcal{F}$ *for* $s \in S$ *, belongs to* $L^1(\pi)$ *. Then the following three statements are equivalent:*

(a) $$\sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=1}^{n} \big(f(Z_i) - Ef(Z_i)\big) \Big| \longrightarrow 0 \quad P\text{-}a.s.$$

(b) $$\sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=1}^{n} \big(f(Z_i) - Ef(Z_i)\big) \Big| \longrightarrow 0 \quad \textit{in } P\textit{-probability}$$

(c) $$\sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=1}^{n} \big(f(Z_i) - Ef(Z_i)\big) \Big| \longrightarrow 0 \quad \textit{in } P\textit{-mean.}$$

Proof. It follows from Corollary 5 of section 3 in [17]. □

We now provide a proof of Theorem 2.1. As shown in section 4, the method of proof is flexible and admits a generalization to the non-stationary setting as described in (1.4).

Proof of Theorem 2.1: By Lemma 2.3, it is enough to show convergence in P-probability in (2.4). Centering, if necessary, we may and do assume that the elements $f \in \mathcal{F}$ have the π-mean zero. The proof is carried out in two steps as follows.

Step 1. We first use Lemma 2.2 to show that the entropy hypothesis (2.3) implies an entropy result for $\mathcal{F}$ with respect to the coupled block sequence $\{ Y_i \}_{i\geq 1}$. We start as follows.

Definition. Let $\hat{N}^X_{b_n}(\varepsilon,\mathcal{F})$ denote the smallest number of open balls in the sup-metric of radius $\varepsilon > 0$ which form a covering of the set of all vectors in $\mathbf{R}^{b_n}$ with coordinates $f(X_i)$ for $i = 1, 2w_n{+}1, 4w_n{+}1 \ldots, (2b_n{-}2)w_n{+}1$ formed by $f \in \mathcal{F}$. Define $\hat{N}^Y_{b_n}(\varepsilon,\mathcal{F})$ in a similar way by replacing X_i with Y_i .

We now show that the entropy condition:

$$\lim_{n\to\infty} \Big(\frac{1}{b_n} \cdot E \log \hat{N}^X_{b_n}(\varepsilon,\mathcal{F}) \Big) = 0 \tag{2.7}$$

is equivalent to the following analogous condition for the coupled block sequence $\{ Y_i \}_{i\geq 1}$:

$$\lim_{n\to\infty} \Big(\frac{1}{b_n} \cdot E \log \hat{N}^Y_{b_n}(\varepsilon,\mathcal{F}) \Big) = 0 \tag{2.8}$$

with $\varepsilon > 0$ being given and fixed.

To verify that these are indeed equivalent entropy conditions, notice that for all $n \geq 1$ we have:

$$\frac{1}{b_n} \cdot \log \hat{N}^Z_{b_n}(\varepsilon,\mathcal{F}) \leq \frac{1}{b_n} \cdot \log \big(C/\varepsilon\big)^{b_n} = \log \big(C/\varepsilon\big)$$

where Z equals X or Y , respectively. Therefore with $n \geq 1$ fixed, there exists a bounded function $g : S^{b_n} \to \mathbf{R}$ such that:

$$\begin{aligned} &Eg\,(Z_1, Z_{2w_n+1}, Z_{4w_n+1}, \ldots, Z_{(2b_n-2)\,w_n+1}) \\ &\quad = \frac{1}{b_n} \cdot E \log \hat{N}^Z_{b_n}(\varepsilon,\mathcal{F}) \end{aligned}$$

where Z equals X or Y , respectively. Moreover $\|g\|_\infty \leq \log \big(C/\varepsilon\big)$, and thus by (2.6) and (2.2') we obtain:

$$\Big| \frac{1}{b_n} \cdot E \log \hat{N}^X_{b_n}(\varepsilon, \mathcal{F}) - \frac{1}{b_n} \cdot E \log \hat{N}^Y_{b_n}(\varepsilon, \mathcal{F}) \Big| \le (b_n - 1) \cdot \beta_{w_n} \cdot \log \big(C/\varepsilon \big) \to 0$$

as $n \to \infty$. This shows the desired equivalence of (2.7) and (2.8).

Moreover, we note that (2.3) trivially implies (2.7), and therefore *the entropy condition (2.3) implies the entropy condition (2.8).* We will use this heavily in the next step.

Step 2. In this step we use lemma 2.2 and condition (2.8) to show that the discrepancy $\sup_{f \in \mathcal{F}} | n^{-1} \sum_{i=1}^n f(X_i) |$ becomes small as n increases.

Indeed, note that we have:

$$\begin{aligned} & P\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^n f(X_i) \Big| > \varepsilon \Big\} \\ & \le P\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{2b_n w_n} f(X_i) \Big| > \varepsilon/2 \Big\} + P\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=2b_n w_n + 1}^{n} f(X_i) \Big| > \varepsilon/2 \Big\} \\ & = P\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{2b_n w_n} f(X_i) \Big| > \varepsilon/2 \Big\} + o(1) \end{aligned}$$

for all $\varepsilon > 0$, and all $n \ge 1$. For the last equality above we use (2.2') from which we obtain:

$$\sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=2b_n w_n + 1}^{n} f(X_i) \Big| \le C \Big(\frac{n - 2b_n w_n}{n} \Big) \to 0$$

as $n \to \infty$. Hence by stationarity and decoupling (Lemma 2.2), we obtain:

(2.9)
$$\begin{aligned} & P\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^n f(X_i) \Big| > \varepsilon \Big\} \\ & \le 2\, P\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \sum_{i \in B_j} f(X_i) \Big| > \varepsilon/4 \Big\} + o(1) \\ & \le 2\, Q\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \sum_{i \in B_j} f(Y_i) \Big| > \varepsilon/4 \Big\} + o(1) \end{aligned}$$

for all $\varepsilon > 0$ and all $n \ge 1$, since $(b_n - 1)\, \beta_{w_n} = o(1)$ by (2.2').

To conclude, it suffices to show that the last term in (2.9) becomes small as n increases. Since the random variables $\sum_{i \in B_j} f(Y_i)$ are independent (and identically distributed) for $1 \le j \le b_n$, it is enough to show that the symmetrized version of the last term in (2.9) becomes arbitrarily small when n increases. Thus by standard symmetrization lemmas it is enough to show that for $\varepsilon > 0$ given and fixed, there exists $n_\varepsilon \ge 1$ such that:

(2.10)
$$(Q \otimes Q_\varepsilon)\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i \in B_j} f(Y_i) \Big| > \varepsilon \Big\} \le \varepsilon$$

for all $n \ge n_\varepsilon$, where $\{ \varepsilon_j \}_{j \ge 1}$ is a Rademacher sequence defined on a probability space $(\Lambda_\varepsilon, \mathcal{B}_\varepsilon, Q_\varepsilon)$ and understood to be independent of the sequence $\{ Y_i \}_{i\geq 1}$, and therefore of the sequence $\{ \sum_{i\in B_j} f(Y_i) \}_{j\geq 1}$ as well.

Note that from Markov's inequality and definition of the coupled sequence $\{ Y_i \}_{i\geq 1}$ we get:

$$
\begin{aligned}
&(Q\otimes Q_\varepsilon)\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i\in B_j} f(Y_i) \Big| > \varepsilon \Big\} \\
&\leq \frac{1}{\varepsilon} E_{Q\otimes Q_\varepsilon} \Big(\sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i\in B_j} f(Y_i) \Big| \Big) \\
&\leq \frac{1}{\varepsilon\, n} \cdot w_n\, E_{Q\otimes Q_\varepsilon} \Big(\sup_{f\in\mathcal{F}} \Big| \sum_{j=1}^{b_n} \varepsilon_j f(Y_{(2j-2)w_n+1}) \Big| \Big) \\
&\leq \frac{1}{2\varepsilon} E_{Q\otimes Q_\varepsilon} \Big(\sup_{f\in\mathcal{F}} \Big| \frac{1}{b_n} \sum_{j=1}^{b_n} \varepsilon_j f(Y_{(2j-2)w_n+1}) \Big| \Big)
\end{aligned}
$$

Since convergence in probability for a uniformly bounded sequence of random variables implies convergence in mean, it is enough to show that there exists $n_\varepsilon \geq 1$ such that:

$$
\text{(2.10')} \qquad (Q\otimes Q_\varepsilon)\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{b_n} \sum_{j=1}^{b_n} \varepsilon_j f(Y_{(2j-2)w_n+1}) \Big| > 2\varepsilon \Big\} < 2\varepsilon
$$

for all $n \geq n_\varepsilon$.

To show (2.10'), proceed as in Yukich [32]. Assume without loss of generality that $\mathcal{F}$ has the uniform bound 1 . Let A denote the event:

$$
\sup_{f\in\mathcal{F}} \Big| \frac{1}{b_n} \sum_{j=1}^{b_n} \varepsilon_j f(Y_{(2j-2)w_n+1}) \Big| > 2\varepsilon
$$

with $\varepsilon > 0$ and $n \geq 1$ fixed. Observe that (2.8) implies the existence of $n_\varepsilon \geq 1$ such that:

$$
\begin{aligned}
&E \log \hat{N}^Y_{b_n}(\varepsilon, \mathcal{F}) \leq \varepsilon^4 b_n \\
&\exp\big(b_n \varepsilon^2(\varepsilon - 1/2)\big) \leq \varepsilon/2
\end{aligned}
$$

for all $n \geq n_\varepsilon$ with $\varepsilon < 1/2$.

By the definition of the entropy number $N = \hat{N}^Y_{b_n}(\varepsilon, \mathcal{F})$, there are vectors x_l in $[-1\,,1]^n$ for $1 \leq l \leq N$ with coordinates $x_{l,i}$ for $i = 1, 2w_n+1, 4w_n+1, \ldots, (2b_n - 2)w_n+1$, such that for all $f \in \mathcal{F}$ we have:

$$
\inf_{1\leq l\leq N} \max_i \big| f(Y_i) - x_{l,i} \big| < \varepsilon
$$

where the max runs over all indices $1, 2w_n+1, 4w_n+1, \ldots, (2b_n-2)w_n+1$. By the triangle inequality we have:

$$
\begin{aligned}
\text{(2.11)} \qquad (Q\otimes Q_\varepsilon)(A) \leq (Q\otimes Q_\varepsilon)\Big\{ \sup_{f\in\mathcal{F}} \frac{1}{b_n} \Big| \sum_{j=1}^{b_n} \varepsilon_j\, f(Y_{(2j-2)w_n+1}) \\
- \sum_{j=1}^{b_n} \varepsilon_j x_{l(f),(2j-2)w_n+1} \Big| > \varepsilon \Big\} + (Q\otimes Q_\varepsilon)\Big\{ \max_{1\leq l\leq N} \frac{1}{b_n} \Big| \sum_{j=1}^{b_n} \varepsilon_j x_{l,(2j-2)w_n+1} \Big| > \varepsilon \Big\}
\end{aligned}
$$

where $x_{l(f)}$ denotes the vector with coordinates $x_{l(f),i}$ satisfying:

$$\max_i \left| f(Y_i) - x_{l(f),i} \right| < \varepsilon$$

with the max as above. The first term on the right-hand side of inequality (2.11) is zero, by choice of $x_{l(f)}$. Applying the standard subgaussian inequality to the second term yields:

$$(2.12) \qquad (Q \otimes Q_\varepsilon)\Big\{ \max_{1 \le l \le N} \frac{1}{b_n} \Big| \sum_{j=1}^{b_n} \varepsilon_j \, x_{l,(2j-2)w_n+1} \Big| > \varepsilon \Big\}$$
$$\le 2N \cdot \exp\left(\frac{-b_n^2 \, \varepsilon^2}{2\, b_n}\right) = 2N \cdot \exp\left(\frac{-b_n \, \varepsilon^2}{2}\right) .$$

Note that for all $n \ge n_\varepsilon$, Markov's inequality implies:

$$(2.13) \qquad Q\Big\{ \log N \ \ge \ b_n \varepsilon^3 \Big\} \le \varepsilon \ .$$

Finally, combining (2.11)–(2.13), the left hand side of (2.10') becomes:

$$(Q \otimes Q_\varepsilon)(A) = \int_A 1_{\{N \ge \exp(b_n \varepsilon^3)\}} d(Q \otimes Q_\varepsilon) + \int_A 1_{\{N < \exp(b_n \varepsilon^3)\}} d(Q \otimes Q_\varepsilon)$$
$$\le \varepsilon + 2 \cdot \exp\left(b_n \varepsilon^3\right) \cdot \exp\left(\frac{-b_n \varepsilon^2}{2}\right) \le 2\varepsilon$$

for all $n \ge n_\varepsilon$. This completes Step 2 and the proof of Theorem 2.1. □

In the remainder of this section we extend Theorem 2.1 to the unbounded case. Since this approach follows in a straightforward way along the lines of Giné and Zinn [9], we will not provide all details.

It is assumed in Theorem 2.1 that the elements $f \in \mathcal{F}$ satisfy $\| f \|_\infty \le C$. To handle the more general case, assume that the envelope $F_{\mathcal{F}}(s) = sup_{f \in \mathcal{F}} | f(s) |$ of $\mathcal{F}$ for $s \in S$, belongs to $L^1(\pi)$, where π is the law of X_1 . Given $R > 0$, define the truncated versions of elements of $\mathcal{F}$ by:

$$f_R(s) = f(s) \cdot 1_{\{F_{\mathcal{F}} \le R\}}(s) \quad \text{for} \quad s \in S \ .$$

Let $N_{n,R}(\varepsilon, \mathcal{F})$ denote the cardinality of the minimal set of open balls in the sup-metric of radius $\varepsilon > 0$, which form a covering of the set of vectors in $\mathbf{R}^n$ of the form $\big(f_R(X_1), \ldots, f_R(X_n)\big)$ when f ranges over $\mathcal{F}$, and where $n \ge 1$ is given and fixed. With this notation, we may now state a generalization of Theorem 2.1 as follows.

Theorem 2.4. *Let* $\{ X_i \}_{i \ge 1}$ *be an absolutely regular sequence of random variables satisfying the condition:*

$$w_n^{-1} \beta_{w_n} = o(n^{-1})$$

for some sequence $w_n = o(n)$ *. Let* $\mathcal{F}$ *be class of functions with envelope* $F_{\mathcal{F}} \in L^1(\pi)$ *, where* π *is the law of* X_1 *. If* $\mathcal{F}$ *satisfies the entropy condition:*

$$\lim_{n \to \infty} \ w_n \ \frac{E \log N_{n,R}(\varepsilon, \mathcal{F})}{n} \ = \ 0$$

for all $\varepsilon > 0$ *and all* $R > 0$ *, then* $\mathcal{F}$ *satisfies the uniform strong law of large numbers (2.4).*

Proof. Follow the proof of Theorem 2.1. In Step 2 observe that by Chebyshev's inequality:

$$P\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=1}^{n} f(X_i) \Big| > \varepsilon \Big\} \le P\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=1}^{n} f(X_i)\cdot 1_{\{F_\mathcal{F}\le R\}} \Big| > \varepsilon/2 \Big\}$$
$$+ P\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=1}^{n} f(X_i)\cdot 1_{\{F_\mathcal{F}> R\}} \Big| > \varepsilon/2 \Big\} \le P\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=1}^{n} f_R(X_i) \Big| > \varepsilon/2 \Big\}$$
$$+ \frac{2}{\varepsilon}\, E\big(|F_\mathcal{F}|\cdot 1_{\{F_\mathcal{F}>R\}}\big)$$

for all $R > 0$, and all $n \ge 1$. Letting $n \to \infty$, Theorem 2.1 shows that the first term after the last inequality sign may be made arbitrarily small. Letting $R \to \infty$, it is clear that the hypothesis $F_\mathcal{F} \in L^1(\pi)$ implies that the second term may also be made arbitrarily small. □

There are clearly other ways to extend and generalize Theorem 2.1. The blocking and decoupling techniques described here may also treat the case of stationary sequences of random variables which have a weak dependence structure, but not necessarily a β-mixing structure. This problem appears worthy of consideration.

3. Uniform ergodic theorems for absolutely regular dynamical systems under the VC entropy condition. The aim of this section is to use the results and methods of section 2 to obtain uniform pointwise ergodic theorems, extending the classic results of Birkhoff, see [13].

Throughout, let $(\Omega, \mathcal{A}, \mu)$ be a probability space, and let T be a measure-preserving transformation of Ω . Then $(\Omega, \mathcal{A}, \mu, T)$ is a *measurable dynamical system*. Let $\kappa : \Omega \to S$ be a measurable function, where $(S, \mathcal{S})$ is a measurable space. For every $l \ge 1$ introduce the σ-algebras:

$$\sigma_1^l = \sigma_1^l(\kappa) = \sigma(\kappa, \kappa\circ T^1, \dots, \kappa\circ T^{l-1})$$
$$\sigma_l^\infty = \sigma_l^\infty(\kappa) = \sigma(\kappa\circ T^l, \kappa\circ T^{l+1}, \dots)\ .$$

The (β,κ)-mixing coefficient of T (or the β-mixing coefficient of T through κ) is defined as follows:

$$\beta_k = \beta_k(\kappa) = \sup_{l\ge 1} \int \sup_{A\in\sigma_{k+l}^\infty} |\,\mu(A\,|\,\sigma_1^l) - \mu(A)\,|\; d\mu \tag{3.1}$$

for all $k \ge 1$. The measurable dynamical system $(\Omega, \mathcal{A}, \mu, T)$ is said to be *absolutely regular through* κ or (β,κ)*-mixing*, if $\beta_k \to 0$ as $k \to \infty$.

Notice that the sequence of random variables $\kappa, \kappa\circ T^1, \kappa\circ T^2, \dots$ is stationary and, when $\beta_k \to 0$ as $k \to \infty$, it is also ergodic. Therefore, as noted at the outset of section 2, we have for every $f \in L^1(\pi)$ with π being the law of κ , the usual pointwise ergodic theorem:

$$\Big| \frac{1}{n} \sum_{i=0}^{n-1} f(\kappa \circ T^i) - \int_\Omega f(\kappa(\omega))\, \mu(d\omega) \Big| \longrightarrow 0 \quad \mu\text{-a.s. } \omega \in \Omega$$

as $n \to \infty$.

We wish to extend this ergodic theorem and obtain a uniform ergodic theorem over a class $\mathcal{F}$ of real valued functions on S , and over a class K of factorizations. The class $\mathcal{F}$ is said to satisfy the *uniform ergodic theorem for* T *with respect to the factorization* κ , if we have:

$$\sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} f(\kappa \circ T^i) - \int_\Omega f(\kappa(\omega))\, \mu(d\omega) \Big| \longrightarrow 0 \quad \mu\text{-a.s. } \omega \in \Omega$$

as $n \to \infty$. In this case we write $\mathcal{F} \in UET(\kappa)$. This approach involves conditions on the entropy number $N_n(\varepsilon, \mathcal{F}, \kappa)$ of $\mathcal{F}$ associated with T through the factorization κ .

Here $N_n(\varepsilon, \mathcal{F}, \kappa)$ denotes the cardinality of the minimal set of open balls in the sup-metric of radius $\varepsilon > 0$ which form a covering of the set of all vectors in $\mathbf{R}^n$ of the form $\big(f(\kappa), f(\kappa \circ T), \ldots, f(\kappa \circ T^{n-1})\big)$ formed by $f \in \mathcal{F}$, where $n \geq 1$ is given and fixed.

Our first result shows that a weighted VC entropy condition insures that $\mathcal{F} \in UET(\kappa)$. We are unaware of a similar result.

Theorem 3.1. *Let* $(\Omega, \mathcal{A}, \mu, T)$ *be an absolutely regular measurable dynamical system through a factorization* $\kappa : \Omega \to S$ *satisfying:*

$$w_n^{-1} \beta_{w_n} = o(n^{-1})$$

for some sequence $w_n = o(n)$. *If* $\mathcal{F}$ *is a uniformly bounded class of functions on* S *satisfying:*

$$\lim_{n \to \infty} w_n \frac{E \log N_n(\varepsilon, \mathcal{F}, \kappa)}{n} = 0$$

for all $\varepsilon > 0$ *, then* $\mathcal{F} \in UET(\kappa)$.

Proof. This follows exactly from Theorem 2.1 upon identifying the random variables X_i with $\kappa \circ T^{i-1}$ for $i \geq 1$. □

It follows trivially from Theorem 2.4 that Theorem 3.1 admits an extension to the case of unbounded $\mathcal{F}$. We will not pursue this. Instead, we consider extensions of Theorem 3.1 to measurable dynamical systems equipped with a family K of factorizations $\kappa : \Omega \to S$. Our main result in this direction may be stated as follows.

Theorem 3.2. *Let* $(\Omega, \mathcal{A}, \mu, T)$ *be a measurable dynamical system, let* $(S, \mathcal{S})$ *be a measurable space, and let* $\mathrm{K} = \{\, \kappa : \Omega \to S \,\}$ *be a family of measurable functions* (*factorizations*) *satisfying:*

(3.2) $$\sup_{\kappa \in \mathrm{K}} w_n^{-1} \beta_{w_n}(\kappa) = o(n^{-1})$$

for some sequence $w_n = o(n)$. *If* $\mathcal{F}$ *is a uniformly bounded class of functions on* S *satisfying the uniform weighted entropy condition:*

$$(3.3) \qquad \lim_{n\to\infty} \sup_{\kappa\in K} w_n \frac{E\log N_n(\varepsilon,\mathcal{F},\kappa)}{n} = 0$$

for all $\varepsilon > 0$ *, then:*

$$(3.4) \qquad \sup_{\kappa\in K} \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=0}^{n-1} f(\kappa\circ T^i) - \int_\Omega f(\kappa(\omega))\,\mu(d\omega) \Big| > \varepsilon \Big\} \longrightarrow 0$$

$$(3.5) \qquad \sup_{\kappa\in K} E \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=0}^{n-1} f(\kappa\circ T^i) - \int_\Omega f(\kappa(\omega))\,\mu(d\omega) \Big| \longrightarrow 0$$

for all $\varepsilon > 0$ *, as* $n \to \infty$ *.*

Proof. The proof is essentially a modification of the proof of Theorem 2.1.

First, given T , construct an associated coupled block sequence of random variables $\{ \upsilon_i \}_{i\geq 1}$ on a probability space $(\Lambda,\mathcal{B},\nu)$ with values in Ω and with the property:

$$\mathcal{L}(\upsilon_1,\ldots,\upsilon_{b_n w_n}) = \bigotimes_1^{b_n} \mathcal{L}(T^0,T^1,\ldots,T^{w_n-1})$$

where $b_n = [n/2w_n]$ for $n \geq 1$. Next, given $\kappa\in K$, write:

$$X_i^\kappa = \kappa\circ T^{i-1} \quad \text{and} \quad Y_i^\kappa = \kappa\circ\upsilon_i \ .$$

for all $i \geq 1$. Then we evidently have:

$$\mathcal{L}(Y_1^\kappa,\ldots,Y_{b_n w_n}^\kappa) = \bigotimes_1^{b_n} \mathcal{L}(X_1^\kappa,\ldots,X_{w_n}^\kappa)$$

for all $n \geq 1$. Following the argument in the proof of Theorem 2.1, we obtain the decoupled inequality:

$$\sup_{\kappa\in K} \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=1}^{n} \Big(f(X_i^\kappa) - \int_\Omega f(X_i^\kappa(\omega))\,\mu(d\omega)\Big) \Big| > \varepsilon \Big\}$$

$$(3.6) \qquad \leq 2 \sup_{\kappa\in K} \nu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{j=1}^{b_n}\sum_{i\in B_j} \Big(f(Y_i^\kappa) - \int_\Lambda f(Y_i^\kappa(\lambda))\,\nu(d\lambda)\Big) \Big| > \varepsilon/4 \Big\}$$

$$+\, o(1) + \sup_{\kappa\in K} (b_n - 1)\beta_{w_n}(\kappa)$$

for all $n \geq 1$. The last term in (3.6) is clearly $o(1)$ by hypothesis (3.2). The first term in (3.6) converges to zero as $n \to \infty$ by the methods of the proof of Theorem 2.1, together with the uniform entropy hypothesis (3.3), and the fact that the centering terms drop out when we randomize. This completes the proof of (3.4).

Finally (3.5) follows by the integration by parts formula $EW = \int_0^\infty \mu\{ W > t \}\,dt$ for the expectation of the random variable $W = \sup_{f\in\mathcal{F}} \big| \frac{1}{n}\sum_{i=0}^{n-1} f(\kappa\circ T^i) - \int_\Omega f(\kappa(\omega))\,\mu(d\omega) \big|$, together with Lebesgue's dominated convergence theorem. □

In the remainder of this section we extend the basic Theorem 2.1 to one-parameter semi-flows $\{ T_t \}_{t\geq 0}$ defined on the probability space $(\Omega,\mathcal{A},\mu)$. By a one-parameter semi-flow, we mean a group of measurable transformations $T_t : \Omega \to \Omega$ with $T_0 =$

identity, and $T_{s+t} = T_s \circ T_t$ for all $s, t \geq 0$. The semi-flow $\{ T_t \}_{t \geq 0}$ is called measurable if the map $(\omega, t) \mapsto T_t(\omega)$ from $\Omega \times [0, \infty)$ into Ω is $\mu \otimes \lambda$-measurable; see [13] for details. Henceforth we will assume that the semi-flow $\{ T_t \}_{t \geq 0}$ is *measure-preserving*, that is, each T_t is measure-preserving for $t \geq 0$.

As above, let $(S, \mathcal{S})$ be a measurable space, let $\kappa : \Omega \to S$ be a measurable function, and let $\mathcal{F}$ be a class of real valued measurable functions defined on S . The class $\mathcal{F}$ is said to satisfy the *uniform ergodic theorem for* $\{ T_t \}_{t \geq 0}$ *with respect to the factorization* κ , whenever:

$$\sup_{f \in \mathcal{F}} \Big| \frac{1}{Z} \int_0^Z f(\kappa \circ T_t)\, dt - \int_\Omega f(\kappa(\omega))\, \mu(d\omega) \Big| \longrightarrow 0 \;\; \mu\text{-a.s. } \omega \in \Omega \tag{3.7}$$

as $Z \to \infty$. In order to apply the above results, we will assume here and henceforth and without further mention that κ satisfies the following regularity condition:

$$\kappa(\omega') = \kappa(\omega'') \Rightarrow \int_0^1 f\big(\kappa \circ T_t(\omega')\big)\, dt = \int_0^1 f\big(\kappa \circ T_t(\omega'')\big)\, dt \tag{3.8}$$

whenever $\omega', \omega'' \in \Omega$ and $f \in \mathcal{F}$.

Under assumption (3.8) we define a measurable map $F : S \times \mathcal{F} \to \mathbf{R}$ satisfying:

$$F(\kappa(\omega), f) = \int_0^1 f\big(\kappa \circ T_t(\omega)\big)\, dt \ .$$

Following the previous definitions, let $N_n(\varepsilon, \mathcal{F}, \kappa)$ denote the smallest number of open balls in the sup-metric of radius $\varepsilon > 0$ which form a covering of the subset of $\mathbf{R}^n$ of the form $\big(F(\kappa, f), F(\kappa \circ T_1, f), \ldots, F(\kappa \circ T_{n-1}, f)\big)$ formed by $f \in \mathcal{F}$, and where $n \geq 1$ is given and fixed. The numbers $N_n(\varepsilon, \mathcal{F}, \kappa)$ are called the entropy numbers of f associated with $\{ T_t \}_{t \geq 0}$ through the factorization κ .

Also, putting $T := T_1$, the (β, κ)-mixing coefficient $\beta_k := \beta_k(\kappa)$ of T for $k \geq 1$ is defined as in (3.1). The semi-flow $\{ T_t \}_{t \geq 0}$ is said to be (β, κ)-mixing, if $\beta_k := \beta_k(\kappa) \to 0$ as $k \to \infty$.

We may now state a uniform ergodic theorem for flows. It appears to be the first of its kind.

Theorem 3.3. *Let* $\{ T_t \}_{t \geq 0}$ *be a measurable measure-preserving semi-flow of the probability space* $(\Omega, \mathcal{A}, \mu)$, $(S, \mathcal{S})$ *a measurable space,* $\kappa : \Omega \to S$ *a measurable function, and* $\mathcal{F}$ *a uniformly bounded class of functions on* S *. Suppose that the semi-flow* $\{ T_t \}_{t \geq 0}$ *satisfies the mixing condition:*

$$w_n^{-1} \beta_{w_n}(\kappa) = o(n^{-1}) \tag{3.9}$$

for some sequence $w_n = o(n)$ *. If* $\mathcal{F}$ *satisfies the weighted entropy condition:*

$$\lim_{n \to \infty} w_n \frac{E \log N_n(\varepsilon, \mathcal{F}, \kappa)}{n} = 0$$

for all $\varepsilon > 0$ *, then* $\mathcal{F}$ *satisifies the uniform ergodic theorem (3.7).*

Proof. This follows from Theorem 3.1 together with the following two facts:

$$\frac{1}{N} \int_0^{N-1} f\big(\kappa \circ T_t(\omega)\big)\, dt = \frac{1}{N} \sum_{i=0}^{N-1} F(\kappa \circ T^i(\omega), f) \quad \text{for all } \omega \in \Omega \tag{3.10}$$

$$\text{(3.11)} \qquad \sup_{f\in\mathcal{F}} \Big| \frac{1}{N}\int_0^{N-1} f\big(\kappa\circ T_t(\omega)\big)\, dt - \frac{1}{Z}\int_0^{Z} f\big(\kappa\circ T_t(\omega)\big)\, dt \Big| \longrightarrow 0 \quad \text{for all } \omega\in\Omega$$

as $N := [Z] \to \infty$. □

It is clear that Theorem 3.3 admits an extension to the case of unbounded $\mathcal{F}$ having an envelope belonging to $L^1(\mu)$. We will not pursue this, but instead consider a generalization of Theorem 3.3 which holds uniformly over a class K of factorizations. The result may be stated as follows.

Theorem 3.4. *Let* $\{ T_t \}_{t\geq 0}$ *be a measurable measure-preserving semi-flow of the probability space* $(\Omega, \mathcal{A}, \mu)$, $(S, \mathcal{S})$ *a measurable space, and* K *a family of factorizations* $\kappa : \Omega \to S$. *Suppose that the semi-flow* $\{ T_t \}_{t\geq 0}$ *satisfies the mixing condition (3.9) uniformly over* K *as follows:*

$$\sup_{\kappa\in\mathrm{K}} w_n^{-1}\beta_{w_n}(\kappa) = o(n^{-1})$$

for some sequence $w_n = o(n)$. *If the uniformly bounded class* $\mathcal{F}$ *of functions on* S *satisfies the weighted entropy condition:*

$$\lim_{n\to\infty} \sup_{\kappa\in\mathrm{K}} w_n \frac{E \log N_n(\varepsilon, \mathcal{F}, \kappa)}{n} = 0$$

for all $\varepsilon > 0$, *then we have:*

$$\sup_{\kappa\in\mathrm{K}} \mu\Big\{ \omega\in\Omega \mid \sup_{f\in\mathcal{F}} \Big| \frac{1}{Z}\int_0^Z f(\kappa\circ T_t(\omega))\, dt - \int_\Omega f(\kappa(\omega))\, \mu(d\omega) \Big| > \varepsilon \Big\} \longrightarrow 0$$

$$\sup_{\kappa\in\mathrm{K}} \int_\Omega \sup_{f\in\mathcal{F}} \Big| \frac{1}{Z}\int_0^Z f(\kappa\circ T_t(\omega))\, dt - \int_\Omega f(\kappa(\omega))\, \mu(d\omega) \Big|\, \mu(d\omega) \longrightarrow 0$$

as $Z \to \infty$.

Proof. This follows along the lines of Theorem 3.3 using the uniform approach of Theorem 3.2. We also make use of (3.10) and a uniformized version of (3.11) as follows:

$$\sup_{\kappa\in\mathrm{K}} \int_\Omega \sup_{f\in\mathcal{F}} \Big| \frac{1}{N}\int_0^{N-1} f\big(\kappa\circ T_t(\omega)\big)\, dt - \frac{1}{Z}\int_0^Z f\big(\kappa\circ T_t(\omega)\big)\, dt \Big|\, \mu(d\omega) \longrightarrow 0$$

as $N := [Z] \to \infty$. □

Example 3.5. We show how Theorem 3.2 applies to moving averages. Let $(\Omega, \mathcal{A}, \mu, T)$ be $(\mathbf{R}^{\mathbf{N}}, \mathcal{B}(\mathbf{R})^{\mathbf{N}}, \mu, \theta)$ where θ denotes the unilateral shift transformation. Let $X_i : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}$ denote the projection onto the i-th coordinate for all $i \geq 1$. Then $\{ X_i \}_{i\geq 1}$ is a stationary sequence of random variables with distribution law μ in $(\mathbf{R}^{\mathbf{N}}, \mathcal{B}(\mathbf{R})^{\mathbf{N}})$. Let K be the family $\{\kappa_m\}_{m\geq 1}$ where $\kappa_m(s_1, s_2, \dots) = s_1 + \ldots + s_m$ for $m \geq 1$. Suppose that $\mathcal{F}$ is a uniformly bounded family of functions from $\mathbf{R}$ into $\mathbf{R}$ satisfying the condition:

$$\lim_{n\to\infty} \sup_{m\geq 1} w_n \frac{E \log N_n(\varepsilon, \mathcal{F}, \kappa_m)}{n} = 0$$

for all $\varepsilon > 0$, where the sequence $w_n = o(n)$ satisfies the uniform mixing rate (3.2). Then it follows from (3.5) that we have:

$$\sup_{m\geq 1} E \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=0}^{n-1} f(X_{i+1}+\ldots+X_{i+m}) - \int_\Omega f(X_1+\ldots+X_m)\, d\mu \Big| \longrightarrow 0$$

as $n \to \infty$. For example, we may take $\mathcal{F}$ to be any family of VC functions.

4. Uniform ergodic theorems for operators under the VC entropy condition.

In this section we obtain uniform ergodic theorems for operators as described in the introduction. It will become clear that the results of this section go beyond those given earlier. In the process we will see that a convergence in probability version of Theorem 2.1 actually holds for sequences of random variables which are neither identically distributed nor stationary.

Throughout, let $(\Omega, \mathcal{A}, \mu)$ denote a probability space, and T a linear operator in $L^1(\mu)$. For $g \in L^1(\mu)$, let $T^i(g)(\omega) := \big(T^i(g)\big)(\omega)$ for all $\omega \in \Omega$. Given $g \in L^1(\mu)$ and a function class $\mathcal{F}$ of maps from $\mathbf{R}$ into $\mathbf{R}$, we wish to find conditions for the uniform convergence:

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=0}^{n-1} \Big(f\big(T^i(g)(\omega)\big) - \int_\Omega f\big(T^i(g)(\omega)\big)\ \mu(d\omega)\Big) \Big| = 0 \tag{4.1}$$

in μ-probability, as $n \to \infty$. This result may be interpreted as a pointwise uniform ergodic theorem for the operator T . To the best of our knowledge, this sort of uniform ergodic theorem has not been studied previously.

We note that if the operator T is induced by means of a measure-preserving transformation, then (4.1) reduces to the setting considered in Theorem 3.1 above. More precisely, letting the operator T be the composition with a measure-preserving transformation τ of Ω , namely $\big(Tg\big)(\omega) = g\big(\tau(\omega)\big)$ for $\omega \in \Omega$, we may recover our previous results. In this way the results of this section generalize and extend Theorem 2.1, Theorem 3.1 and Theorem 3.2.

Before stating the main result we introduce some notation. Let $g \in L^1(\mu)$ be fixed. For every $l \geq 1$ introduce the σ-algebras:

$$\sigma_1^l = \sigma_1^l(g) = \sigma(g, T^1(g), \ldots, T^{l-1}(g))$$

$$\sigma_l^\infty = \sigma_l^\infty(g) = \sigma(T^l(g), T^{l+1}(g), \ \ldots\ ) \ .$$

The β-mixing coefficient for the operator T with respect to g is defined as follows:

$$\beta_k = \beta_k(g) = \sup_{l\geq 1} \int \sup_{A\in\sigma_{k+l}^\infty} |\, \mu(A \,|\, \sigma_1^l) - \mu(A)\,|\ d\mu$$

for all $k \geq 1$. The measure space $(\Omega, \mathcal{A}, \mu)$ together with the operator T is said to be (β, g)-*mixing*, if $\beta_k \to 0$ as $k \to \infty$. Finally, the class $\mathcal{F}$ is said to satisfy the *uniform ergodic theorem for* T *with respect to* g , if (4.1) holds. In this case we write $\mathcal{F} \in UET(g)$.

It turns out that the methods employed in section 2, which hold for stationary sequences of random variables, may be generalized to treat the non-stationary case. In this way we will find sufficient conditions for $\mathcal{F} \in UET(g)$. As before, the approach involves conditions on the entropy number $N_n(\varepsilon, \mathcal{F}, g)$ of $\mathcal{F}$ with respect to T and g .

Here $N_n(\varepsilon, \mathcal{F}, g)$ denotes the cardinality of the minimal set of open balls in the sup-metric of radius $\varepsilon > 0$ which form a covering of the set of all vectors in $\mathbf{R}^n$ of the form $\big(f(g), f(T(g)), \ldots, f(T^{n-1}(g))\big)$ where f ranges over $\mathcal{F}$, and where $n \geq 1$ is given and fixed.

The main result shows that a weighted VC entropy condition implies that $\mathcal{F} \in UET(g)$.

Theorem 4.1. *Suppose that the measure space* $(\Omega, \mathcal{A}, \mu)$ *and the operator* T *in* $L^1(\mu)$ *are* (β, g)*-mixing, where* $g \in L^1(\mu)$ *is fixed. Suppose that the* β*-mixing coefficients for the operator* T *with respect to* g *satisfy:*

$$\beta_{w_n} = o(n^{-1}) \tag{4.2}$$

for some sequence $w_n = o(n)$ *. If* $\mathcal{F}$ *is a uniformly bounded class of functions on* $\mathbf{R}$ *satisfying:*

$$\lim_{n\to\infty} w_n \, \frac{E \log N_n(\varepsilon, \mathcal{F}, g)}{n} = 0 \tag{4.3}$$

for all $\varepsilon > 0$ *, then* $\mathcal{F} \in UET(g)$ *.*

Remark. As the following proof shows, T can be a non-linear operator, that is, an arbitrary map from $L^1(\mu)$ into $L^1(\mu)$.

Proof. As noted already, the random variables:

$$X_1 = g \, , X_2 = T(g) \, , X_3 = T^2(g) \, , \ldots$$

do not form a stationary sequence, so Theorem 4.1 is not immediate from Theorem 2.1. Additionally, it does not seem possible to apply Lemma 2.3 to deduce a.s.-convergence. Nonetheless, we may prove convergence to zero in μ-probability by adapting the methods used to prove Theorem 2.1. This is done as follows.

First, notice that Eberlein's lemma 2.2 holds for sequences of random variables which are not identically distributed. Therefore, letting $\{ Y_i \}_{i\geq 1}$ be a sequence of random variables defined on a probability space $(\Lambda, \mathcal{B}, \nu)$ with independent blocks satisfying:

$$\mathcal{L}(Y_1, \ldots, Y_{2b_n w_n}) = \mathcal{L}\big(g, T(g), \ldots, T^{w_n-1}(g)\big) \otimes \mathcal{L}\big(T^{w_n}(g), T^{w_n+1}(g), \ldots, T^{2w_n-1}(g)\big)$$
$$\otimes \ldots \otimes \mathcal{L}\big(T^{(2b_n-1)w_n}(g), T^{(2b_n-1)w_n+1}(g), \ldots, T^{2b_n w_n-1}(g)\big)$$

with $b_n = [n/2w_n]$ for $n \geq 1$, we may modify the proof of Step 1 of Theorem 2.1 as follows.

Definition. Let $\hat{N}^X_{b_n w_n}(\varepsilon, \mathcal{F})$ denote the smallest number of open balls in the sup-metric of radius $\varepsilon > 0$ which form a covering of the set of all vectors in $\mathbf{R}^{b_n w_n}$ with coordinates $f(X_i)$ for $i = 1, \ldots, w_n, 2w_n{+}1, \ldots, 3w_n, \ \ldots \ , (2b_n{-}2)w_n{+}1, \ldots, (2b_n{-}1)w_n$ formed by $f \in \mathcal{F}$. Define $\hat{N}^Y_{b_n w_n}(\varepsilon, \mathcal{F})$ in a similar way by replacing X_i with Y_i .

We now show that the entropy condition:

$$\lim_{n\to\infty} \left(\frac{w_n}{n} \cdot E \log \hat{N}^X_{b_n w_n}(\varepsilon, \mathcal{F}) \right) = 0 \tag{4.4}$$

is equivalent to the following analogous condition for the coupled block sequence $\{ Y_i \}_{i\geq 1}$:

$$\lim_{n\to\infty} \left(\frac{w_n}{n} \cdot E \log \hat{N}^Y_{b_n w_n}(\varepsilon, \mathcal{F}) \right) = 0 \tag{4.5}$$

with $\varepsilon > 0$ being given and fixed.

To verify that these are indeed equivalent entropy conditions, notice that for all $n \geq 1$ we have:

$$\frac{w_n}{n} \cdot \log N^Z_n(\varepsilon, \mathcal{F}) \leq \frac{w_n}{n} \cdot \log \big(C/\varepsilon\big)^{b_n w_n} \leq w_n \cdot \log \big(C/\varepsilon\big)^{1/2}$$

where Z equals X or Y , respectively. Therefore with $n \geq 1$ fixed, there exists a bounded function $g : S^{b_n w_n} \to \mathbf{R}$ such that:

$$Eg\,(Z_1, \ldots, Z_{w_n}, Z_{2w_n+1}, \ldots, Z_{3w_n}, \ \ldots \ , Z_{(2b_n-2)w_n+1}, \ldots, Z_{(2b_n-1)w_n})$$
$$= \frac{w_n}{n} \cdot E \log \hat{N}^Z_{b_n w_n}(\varepsilon, \mathcal{F})$$

where Z equals X or Y , respectively. Moreover $\|g\|_\infty \leq w_n \cdot \log \big(C/\varepsilon\big)^{1/2}$, and thus by (2.6) and (4.2) we obtain:

$$\Big| \frac{w_n}{n} \cdot E \log \hat{N}^X_{b_n w_n}(\varepsilon, \mathcal{F}) - \frac{w_n}{n} \cdot E \log \hat{N}^Y_{b_n w_n}(\varepsilon, \mathcal{F}) \Big| \leq (b_n - 1) \cdot \beta_{w_n} \cdot w_n \cdot \log(C/\varepsilon)^{1/2}$$
$$\leq n\, \beta_{w_n} \cdot \log(C/\varepsilon)^{1/2} \to 0$$

as $n \to \infty$. This shows the desired equivalence of (4.4) and (4.5).

Moreover, we note that (4.3) trivially implies (4.4), and therefore *the entropy condition (4.3) implies the entropy condition (4.5).* We will use this heavily in the next step.

Concerning Step 2 of the proof of Theorem 2.1, we need to make the following modifications to the decoupling arguments:

$$\mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{2b_n w_n - 1} \Big(f\big(T^i(g)\big) - \int_\Omega f\big(T^i(g)(\omega)\big)\ \mu(d\omega) \Big) \Big| > \varepsilon/2 \Big\}$$
$$\leq \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \sum_{i\in B_j} \Big(f\big(T^{i-1}(g)\big) - \int_\Omega f\big(T^{i-1}(g)(\omega)\big)\ \mu(d\omega) \Big) \Big| > \varepsilon/4 \Big\}$$
$$+\ \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \sum_{i\in \hat{B}_j} \Big(f\big(T^{i-1}(g)\big) - \int_\Omega f\big(T^{i-1}(g)(\omega)\big)\ \mu(d\omega) \Big) \Big| > \varepsilon/4 \Big\}$$
$$\leq \nu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \sum_{i\in B_j} \big(f(Y_i) - Ef(Y_i) \big) \Big| > \varepsilon/4 \Big\} \tag{4.6}$$
$$+\ \nu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \sum_{i\in \hat{B}_j} \big(f(Y_i) - Ef(Y_i) \big) \Big| > \varepsilon/4 \Big\} + 2(b_n - 1)\, \beta_{w_n}$$

for all $n \geq 1$, where the last inequality follows by Eberlein's lemma 2.2. Clearly, as B_j and $\hat{B}_j$ play symmetric roles, the first two terms in (4.6) have an identical form and it suffices to bound the first term by ε. Since the random variables $\sum_{i \in B_j} (f(Y_i) - Ef(Y_i))$ are independent for $1 \leq j \leq b_n$, by standard symmetrization lemmas it is enough to show that for $\varepsilon > 0$ given and fixed, there exists $n_\varepsilon \geq 1$ such that:

$$(4.7) \qquad (\nu \otimes \nu_\varepsilon) \Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i \in B_j} f(Y_i) \Big| > 2\varepsilon \Big\} \leq 2\varepsilon$$

for all $n \geq n_\varepsilon$, where $\{ \varepsilon_j \}_{j \geq 1}$ is a Rademacher sequence defined on a probability space $(\Lambda_\varepsilon, \mathcal{B}_\varepsilon, \nu_\varepsilon)$ and understood to be independent of the sequence $\{ Y_i \}_{i \geq 1}$, and therefore of the sequence $\{ \sum_{i \in B_j} (f(Y_i) - Ef(Y_i)) \}_{j \geq 1}$ as well.

To show (4.7), proceed as in Step 2 of the proof of Theorem 2.1. Assume without loss of generality that $\mathcal{F}$ has the uniform bound 1. Let A denote the event:

$$\sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i \in B_j} f(Y_i) \Big| > 2\varepsilon$$

with $\varepsilon > 0$ and $n \geq 1$ fixed. Observe that (4.5) implies the existence of $n_\varepsilon \geq 1$ such that:

$$w_n \cdot E \log \hat{N}^Y_{b_n w_n}(\varepsilon, \mathcal{F}) \leq \varepsilon^4 n$$

$$\exp\big(n\varepsilon^2(\varepsilon - 1)/w_n\big) \leq \varepsilon/2$$

for all $n \geq n_\varepsilon$ with $\varepsilon < 1$.

By the definition of the entropy number $N = \hat{N}^Y_{b_n w_n}(\varepsilon, \mathcal{F})$, there are vectors x_l in $[-1, 1]^n$ for $1 \leq l \leq N$ with coordinates $x_{l,i}$ for $i = 1, \ldots, w_n, 2w_n + 1, \ldots, 3w_n, \ \ldots \ , (2b_n - 2)w_n + 1, \ldots, (2b_n - 1)w_n$, such that for all $f \in \mathcal{F}$ we have:

$$\inf_{1 \leq l \leq N} \max_i \big| f(Y_i) - x_{l,i} \big| < \varepsilon$$

where the max runs over all indices $1, \ldots, w_n, 2w_n + 1, \ldots, 3w_n, \ \ldots \ , (2b_n - 2)w_n + 1, \ \ldots \ , (2b_n - 1)w_n$. By the triangle inequality we have:

$$(4.8) \qquad (\nu \otimes \nu_\varepsilon)(A) \leq (\nu \otimes \nu_\varepsilon) \Big\{ \sup_{f \in \mathcal{F}} \frac{1}{n} \Big| \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i \in B_j} f(Y_i) - \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i \in B_j} x_{l(f),i} \Big| > \varepsilon \Big\}$$
$$+ (\nu \otimes \nu_\varepsilon) \Big\{ \max_{1 \leq l \leq N} \frac{1}{n} \Big| \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i \in B_j} x_{l,i} \Big| > \varepsilon \Big\}$$

where $x_{l(f)}$ denotes the vector with coordinates $x_{l(f),i}$ satisfying:

$$\max_i \big| f(Y_i) - x_{l(f),i} \big| < \varepsilon$$

with the max as above. The first term on the right-hand side of inequality (4.8) is zero, by choice of $x_{l(f)}$. Applying the standard subgaussian inequality to the second term yields:

$$(4.9) \qquad (\nu \otimes \nu_\varepsilon) \Big\{ \max_{1 \leq l \leq N} \frac{1}{n} \Big| \sum_{j=1}^{b_n} \varepsilon_j \cdot \sum_{i \in B_j} x_{l,i} \Big| > \varepsilon \Big\}$$

$$\leq 2N \cdot \max_{1 \leq l \leq N} \exp\Big(\frac{-n^2 \varepsilon^2}{2} \Big(\sum_{j=1}^{b_n} \Big(\sum_{i \in B_j} x_{l,i} \Big)^2 \Big)^{-1} \Big)$$

$$\leq 2N \cdot \exp\Big(\frac{-n^2 \varepsilon^2}{2 b_n w_n^2} \Big) \leq 2N \cdot \exp\Big(\frac{-n\, \varepsilon^2}{w_n} \Big) \ .$$

Note that for all $n \geq n_\varepsilon$, Markov's inequality implies:

$$\nu\Big\{ \log N \geq \frac{n \varepsilon^3}{w_n} \Big\} \leq \varepsilon \ . \tag{4.10}$$

Finally, combining (4.8)–(4.10), the left hand side of (4.7) becomes:

$$(\nu \otimes \nu_\varepsilon)(A) = \int_A 1_{\{N \geq \exp(n\varepsilon^3/w_n)\}} d(\nu \otimes \nu_\varepsilon) + \int_A 1_{\{N < \exp(n\varepsilon^3/w_n)\}} d(\nu \otimes \nu_\varepsilon)$$

$$\leq \varepsilon + 2 \cdot \exp\Big(\frac{n\, \varepsilon^3}{w_n} \Big) \cdot \exp\Big(\frac{-n\, \varepsilon^2}{w_n} \Big) \leq 2\varepsilon$$

for all $n \geq n_\varepsilon$. This proves the desired convergence in μ-probability. □

It is clear that Theorem 4.1 admits an extension to the case of unbounded $\mathcal{F}$ under an integrability condition. We will not pursue this. Instead, we consider a generalization of Theorem 4.1 which holds uniformly over a class $\mathcal{G}$ of functions from $L^1(\mu)$. The result may be stated as follows.

Theorem 4.2. *Let T be a linear operator in $L^1(\mu)$, where $(\Omega, \mathcal{A}, \mu)$ is a probability space. Let $\mathcal{G}$ be a family of functions from $L^1(\mu)$ satisfying:*

$$\sup_{g \in \mathcal{G}} \beta_{w_n}(g) = o(n^{-1}) \tag{4.11}$$

for some sequence $w_n = o(n)$. If $\mathcal{F}$ is a uniformly bounded class of functions on $\mathbf{R}$ satisfying the uniform weighted entropy condition:

$$\lim_{n \to \infty} \sup_{g \in \mathcal{G}} w_n \frac{E \log N_n(\varepsilon, \mathcal{F}, g)}{n} = 0 \tag{4.12}$$

for all $\varepsilon > 0$, then:

$$\sup_{g \in \mathcal{G}} \mu\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} \Big(f(T^i(g)) - \int_\Omega f(T^i(g)(\omega))\, \mu(d\omega) \Big) \Big| > \varepsilon \Big\} \longrightarrow 0 \tag{4.13}$$

$$\sup_{g \in \mathcal{G}} E \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} \Big(f(T^i(g)) - \int_\Omega f(T^i(g)(\omega))\, \mu(d\omega) \Big) \Big| \longrightarrow 0 \tag{4.14}$$

for all $\varepsilon > 0$, as $n \to \infty$.

Remark. Again, as the following proof shows, T can be a non-linear operator, that is, an arbitrary map from $L^1(\mu)$ into $L^1(\mu)$.

Proof. This follows along the lines of the proof of Theorem 3.2. The lack of stationarity may be overcome as in the proof of Theorem 4.1. □

In the next example we show how Theorem 4.2 applies to moving averages. In this context we find it convenient to recall Example 3.5. Except for the work of de Acosta and Kuelbs [1], we are unaware of limit theorems for moving (delayed) averages in the Banach space setting.

Example 4.3. Let $(\Omega, \mathcal{A}, \mu)$ be $(\mathbf{R}^{\mathbf{N}}, \mathcal{B}(\mathbf{R})^{\mathbf{N}}, \mu)$, and let θ denote the unilateral shift transformation of $\mathbf{R}^{\mathbf{N}}$. It should be noted that θ is not supposed to be stationary with respect to μ . Let T be the composition operator with θ in $L^1(\mu)$. Let $\mathcal{G}$ be the family $\{ \pi_m \}_{m \geq 1}$, where $\pi_m : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}$ denotes the projection onto the m-th coordinate. Put $X_m(\omega) = T(\pi_m)(\omega)$ for all $\omega \in \Omega$, and all $m \geq 1$. Then $\{ X_m \}_{m \geq 1}$ is a sequence of random variables with distribution law μ in $(\mathbf{R}^{\mathbf{N}}, \mathcal{B}(\mathbf{R})^{\mathbf{N}})$. Suppose that $\mathcal{F}$ is a uniformly bounded family of functions from $\mathbf{R}$ into $\mathbf{R}$ satisfying the condition:

$$\lim_{n \to \infty} \sup_{m \geq 1} w_n \frac{E \log N_n(\varepsilon, \mathcal{F}, \pi_m)}{n} = 0$$

for all $\varepsilon > 0$, where the sequence $w_n = o(n)$ satisfies the uniform mixing rate (4.11). Then it follows from (4.14) that we have:

$$\sup_{m \geq 1} E \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} f(X_{i+m}) - \int_\Omega f(X_{i+m}) \, d\mu \Big| \longrightarrow 0$$

as $n \to \infty$.

It is easily seen that Example 3.5 admits a similar generalization to the non-stationary case. We leave the formulation of this result and the remaining details to the reader.

Acknowledgment. The authors would like to thank Bin Yu for sharing her informative and useful preprints.

Appendix. After the completion of this paper we learned of the recent work of Nobel and Dembo* in which it is shown that the uniform strong law of large numbers (2.4) holds whenever:

$$\sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} \big(f(Z_i) - Ef(Z_i) \big) \Big| \longrightarrow 0 \ \ a.s.$$

where $\{Z_i\}_{i \geq 1}$ denotes an i.i.d. sequence of random variables with common distribution π . In terms of metric entropy, their theorem takes the form:

Theorem A (Nobel and Dembo)* *Let* $\mathcal{F} \subset L^1(\pi)$ *be a uniformly bounded class of functions. If for all* $\varepsilon > 0$ *:*

(A.1) $$\lim_{n \to \infty} \frac{E \log N_n^Z(\varepsilon, \mathcal{F})}{n} = 0$$

then $\mathcal{F}$ *satisfies the uniform SLLN for the absolutely regular sequence* $\{X_i\}_{i \geq 1}$ *:*

(A.2) $$\sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} \big(f(X_i) - Ef(X_i) \big) \Big| \longrightarrow 0 \ \ P\text{-}a.s.$$

* Nobel, A.B. *and* Dembo, A. (1993). A note on uniform laws of averages for dependent processes. *Stat. Probab. Lett.*, to appear.

In other words, the entropy condition (A.1) with respect to the auxiliary sequence $\{Z_i\}_{i\geq 1}$ yields the limit result (A.2) for the absolutely regular sequence $\{X_i\}_{i\geq 1}$. In both theory and practice it seems more desirable to obtain (A.2) under entropy conditions on the *observable* sequence $\{X_i\}_{i\geq 1}$.

Despite this drawback Theorem A allows a partial simplification of the proof of Theorem 2.1. To see this, notice that it suffices to show that the conditions:

(A.3) $$\lim_{n\to\infty} w_n \frac{E \log N_n^X(\varepsilon, \mathcal{F})}{n} = 0 \ , \ \ w_n^{-1} \beta_{w_n} = o(n^{-1}) \ , \ \ w_n = o(n)$$

imply (A.1)**. This implication readily follows using a judicious choice of g in the decoupling estimate (2.6), as shown by proving the equivalence of (2.7) and (2.8). After this proof it is pointed out that (A.3) implies (2.7), and thus (2.8) as well. However, (2.8) is equivalent to (A.1) since the limit in (A.1) always exists (see [28]). Moreover, this shows that under (2.2), the condition (A.1) is equivalent to (2.7). In this way we obtain a refinement of Theorem A, since the entropy condition (2.7) is in terms of the observable sequence $\{X_i\}_{i\geq 1}$.

While this approach clearly leads to a simplification of Theorem 2.1, it must be pointed out that the approach is not applicable to the case of non-i.i.d. random variables $\{X_i\}_{i\geq 1}$. Thus this approach does not simplify the proof of the uniform ergodic theorem for operators (Theorem 4.1). It may be similarly seen that these methods do not simplify the proofs of the uniform theorems (Theorem 3.2 and Theorem 3.4).

Thus, while the proof of Theorem 2.1 is perhaps not the shortest, it has the advantage that it may be modified to treat more general situations which do not seem accessible by known methods. We find it convenient to recall (2.10) for this purpose.

REFERENCES

[1] ACOSTA, A. de *and* KUELBS, J. (1983). Limit theorems for moving averages of independent random vectors. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 64 (67-123).

[2] ARCONES, M. A. *and* YU, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.* 7 (47-71).

[3] BRADLEY, R. C. (1983). Absolute regularity and functions of Markov chains. *Stochastic Precesses Appl.* 14 (67-77).

[4] DEHLING, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 63 (393-432).

[5] DUDLEY, R. M. (1984). A course on empirical processes. *Ecole d'Eté de Probabilitiés de Saint-Flour, XII-1982. Lecture Notes in Math.* 1097, Springer-Verlag Berlin Heidelberg (1-142).

[6] DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann.*

** It is not clear whether the converse implication (A.1) ⇒ (A.3) holds. Indeed, given an entropy bound for $E \log N_n^Z(\varepsilon, \mathcal{F})$ in the i.i.d. case, it is an interesting open question whether this implies a similar bound in the non-i.i.d. case.

Probab. 15 (1306-1326).

[7] EBERLEIN, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Stat. Probab. Lett.* 2 (291-293).

[8] GAENSSLER, P. (1983). Empirical Processes. *IMS Lecture Notes-Monograph Series* 3.

[9] GINÉ, J. *and* ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* 12 (929-989).

[10] HALMOS, P. R. (1956). *Lectures on Ergodic Theory*. The Mathematical Society of Japan.

[11] HOFFMANN-JØRGENSEN, J. (1984). Necessary and sufficient conditions for the uniform law of large numbers. *Proc. Conf. Probab. Banach Spaces V (Medford 1984), Lecture Notes in Math.* 1153, Springer-Verlag Berlin Heidelberg (258-272).

[12] JUNCO, A. del *and* STEELE, J. M. (1977). Moving averages of ergodic processes. *Metrika* 24 (35-43).

[13] KRENGEL, U. (1985). *Ergodic Theorems*. Walter de Gruyter & Co., Berlin.

[14] KUELBS, J. *and* ZINN, J. (1979). Some stability results for vector valued random variables. *Ann. Probab.* 7 (75-84).

[15] MASSART, P. (1988). Invariance principles for empirical processes: the weakly dependent case. *PhD thesis, University of Paris.*

[16] PESKIR, G. (1992). Uniform convergence of reversed martingales. *Math. Inst. Aarhus, Preprint Ser.* No. 21, (27 pp). *J. Theoret. Probab.* 8, 1995 (387-415).

[17] PESKIR, G. *and* WEBER, M. (1992). Necessary and sufficient conditions for the uniform law of large numbers in the stationary case. *Math. Inst. Aarhus, Preprint Ser.* No. 27, (26 pp). *Proc. Funct. Anal.* IV (Dubrovnik 1993), *Various Publ. Ser.* Vol. 43, 1994 (165-190).

[18] PESKIR, G. *and* WEBER, M. (1993). The uniform ergodic theorem for dynamical systems. *Math. Inst. Aarhus, Preprint Ser.* No. 14, (30 pp). *Proc. Convergence in Ergodic Theory and Probability* (Columbus 1993), *Ohio State Univ. Math. Res. Inst. Publ.* Vol. 5, Walter de Gruyter, Berlin, 1996 (305-332).

[19] PETERSEN, K. (1983). *Ergodic Theory*. Cambridge University Press.

[20] PHILIPP, W. (1984). Invariance principles for sums of mixing random elements and the multivariate empirical process. *Limit theorems in probability and statistics, 2nd Colloq., Veszprém/Hung. 1982, Vol. II, Colloq. Math. Soc. János Bolyai* 36 (843-873).

[21] PHILIPP, W. (1986). Invariance principles for independent and weakly dependent random variables. *Dependence in probability and statistics, Conf. Oberwolfach 1985, Prog. Probab. Stat.* 11 (225-268).

[22] POLLARD, D. (1981). Limit theorems for empirical processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 57 (181-185).

[23] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag New York Inc.

[24] POLLARD, D. (1990). Empirical Processes: Theory and Applications. *NSF-CBMS*

Regional Conference Series in Probability in Statistics, Vol. 2.

[25] STEELE, J. M. (1978). Empirical discrepancies and subadditive processes. *Ann. Probab.* 6 (118-127).

[26] TALAGRAND, M. (1987). The Glivenko-Cantelli problem. *Ann. Probab.* 15 (837-870).

[27] VAPNIK, V. N. *and* CHERVONENKIS, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* 16 (264-280).

[28] VAPNIK, V. N. *and* CHERVONENKIS, A. Ya. (1981). Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory Probab. Appl.* 26 (532-553).

[29] VOLKONSKII, V. A. *and* ROZANOV, Yu. A. (1959). Some limit theorems for random functions I. *Theory Probab. Appl.* 4 (178-197).

[30] VOLKONSKII, V. A. *and* ROZANOV, Yu. A. (1961). Some limit theorems for random functions II. *Theory Probab. Appl.* 6 (186-197).

[31] YU, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* 22 (94-116).

[32] YUKICH, J. E. (1985). Sufficient conditions for the uniform convergence of means to their expectations. *Sankhya*, Ser. A 47 (203-208).

[33] YUKICH, J. E. (1986). Rates of convergence for classes of functions: The non-i.i.d. case. *J. Multivariate Anal.* 20 (175-189).

Goran Peskir
Department of Mathematical Sciences
University of Aarhus, Denmark
Ny Munkegade, DK-8000 Aarhus
home.imf.au.dk/goran
goran@imf.au.dk

Joseph Yukich
Department of Mathematics
Lehigh University
Bethlehem PA 18015
USA
joseph.yukich@lehigh.edu