Lecture Notes Series No. 66, 2000, *Dept. Math. Univ. Aarhus*
Math. Inst. Aarhus, Preprint Ser. No. 3, 1994, (118 pp)

From Uniform Laws of Large Numbers to Uniform Ergodic Theorems

GORAN PESKIR

The purpose of these lectures is to present three different approaches with their own methods for establishing uniform laws of large numbers and uniform ergodic theorems for dynamical systems. The presentation follows the principle according to which the i.i.d. case is considered first in great detail, and then attempts are made to extend these results to the case of more general dependence structures. The lectures begin (Chapter 1) with a review and description of classic laws of large numbers and ergodic theorems, their connection and interplay, and their infinite dimensional extensions towards uniform theorems with applications to dynamical systems. The first approach (Chapter 2) is of metric entropy with bracketing which relies upon the Blum-DeHardt law of large numbers and Hoffmann-Jørgensen's extension of it. The result extends to general dynamical systems using the uniform ergodic lemma (or Kingman's subadditive ergodic theorem). In this context metric entropy and majorizing measure type conditions are also considered. The second approach (Chapter 3) is of Vapnik and Chervonenkis. It relies upon Rademacher randomization (subgaussian inequality) and Gaussian randomization (Sudakov's minoration) and offers conditions in terms of random entropy numbers. Absolutely regular dynamical systems are shown to support the VC theory using a blocking technique and Eberlein's lemma. The third approach (Chapter 4) is aimed to cover the wide sense stationary case which is not accessible by the previous two methods. This approach relies upon the spectral representation theorem and offers conditions in terms of the orthogonal stochastic measures which are associated with the underlying dynamical system by means of this theorem. The case of bounded variation is covered, while the case of unbounded variation is left as an open question. The lectures finish with a supplement in which the role of uniform convergence of reversed martingales towards consistency of statistical models is explained via the concept of Hardy's regular convergence.

Mathematics Subject Classification 2000. Primary 28D05, 28D10, 47A35, 60B12. Secondary 60G10, 60G42, 62F10, 82B05.
Key words and phrases: Uniform law of large numbers, the Glivenko-Cantelli theorem, Kolmogorov's (strong) law of large numbers, the VC law of large numbers, the VC class of sets, uniform ergodic theorem for dynamical systems, the maximal ergodic lemma (inequality), the uniform ergodic lemma (inequality), Lioville's theorem, Birkhoff's (pointwise) ergodic theorem, Von Neumann's (mean) ergodic theorem, Kingman's (subadditive) ergodic theorem, Hamiltonian flow, Poincaré's recurrence theorem, metric entropy (with bracketing), the subgaussian inequality, Sudakov's minoration, Rademacher (Gaussian) randomization, Eberlein's lemma, absolutely regular, semi-flow, moving average, the Kakutani-Rokhlin lemma, the uniform mean-square ergodic theorem, wide sense stationary, the spectral representation theorem, orthogonal stochastic measure, almost periodic sequence, Hardy's regular convergence, reversed martingale, consistency of statistical models.

PREFACE

These lecture notes are on the uniform law of large numbers with a view to uniform ergodic theorems for dynamical systems. Our main aim was not to give the final word on the subject, but to motivate its development. As two disjointed subject areas (modern probability theory of empirical processes and classic theory of ergodic theorems) are coming across and interacting in these notes, the reader requires some familiarity with both probability and ergodic theory. In order to aid the reading and understanding of the material presented we have tried to keep technical complexity to a minimum. Occasionally, we needed to refer to other literature sources for the reader to obtain additional information, but despite this fact we believe that the main text is self-contained. We also believe that a reader belonging to either of the subject areas mentioned should have no difficulty in grasping the main points and ideas throughout. Moreover, if we succeed in stimulating further work, our aim will be gladly accomplished.

Aarhus, August 1998 Goran Peskir

Acknowledgments. The author gratefully acknowledges support from *Danish Research Academy* and *Danish Natural Science Research Council*. The author extends his thanks to the *Institut de Recherche Mathematique Avancee* (Universite Louis Pasteur, Strasbourg) for providing support and excellent working conditions during the month of February in 1994, where the main body of the present text has been presented and completed. In this context thanks are particularly due to *M. Weber* and *X. Fernique*. The author further expresses his sincere thanks to *N. Bingham, S. E. Graversen*, *J. Hoffmann-Jørgensen*, *M. Weber* and *J. E. Yukich*.

CONTENTS

1. Introduction

The aim of these lectures is to present solutions to the following problem. Given a dynamical system $(X, \mathcal{A}, \mu, T)$ and a family $\mathcal{F}$ of maps from X into $\mathbf{R}$, find and examine conditions under which the *uniform ergodic theorem*:

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j) - \int_X f \, d\mu \right| \to 0 \quad \mu\text{-a.a.} \tag{1.1}$$

is valid as $n \to \infty$. This problem is fundamental in many ways, and it was only recently realized (see Section 1.3 below) that it has not been studied previously, except in a rather special (i.i.d.) case (the Glivenko-Cantelli theorem), or a very general (operator) case (the Yosida-Kakutani theorem).

1. Our approach towards (1.1) in these lectures is the following. We first consider a particular case of (1.1) where $(X, \mathcal{A}, \mu)$ equals the countable product of a probability space with itself, and where T equals the (unilateral) shift. In this case (1.1) reduces (see Section 1.2 below) to the well-investigated *uniform law of large numbers*:

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j) - Ef(\xi_1) \right| \to 0 \quad \mu\text{-a.s.} \tag{1.2}$$

where $\{\xi_j \mid j \geq 1\}$ is an i.i.d. sequence of random variables and $n \to \infty$. In this context we first present known methods for the uniform law of large numbers in great detail, and then try to extend them towards (1.1) in a case as general as possible.

2. The methods we present in this process are the following. First, we consider the Blum-DeHardt approach (Chapter 2) which uses the concept of *metric entropy with bracketing*. It appeared in the papers of Blum [7] and DeHardt [15], and presently offers the best known sufficient condition for (1.2). A necessary and sufficient condition for (1.2) which involves the Blum-DeHardt condition as a particular case was found by Hoffmann-Jørgensen [39]. It was shown later in [66] and [67] that this result extends to (1.1) in the case of *general* dynamical systems. Second, we consider the Vapnik-Chervonenkis approach (Chapter 3) which uses the concept of *random entropy numbers*. It appeared in the papers of Vapnik and Chervonenkis [89] and [90]. It was shown in [68] that this result extends to (1.1) in the case of *absolutely regular* dynamical systems. Third, since both of the previous two methods could adequately cover only the (strict sense) stationary ergodic case, we make an independent attempt in Chapter 4 to treat the *wide sense stationary* case. Our approach in this context relies upon the spectral representation theorem and offers conditions in terms of the orthogonal stochastic measures. From the point of view of applications, we find it useful to recall that (1.1) and (1.2) play an important role in establishing *consistency of statistical models*. In this context *reversed martingales* happen to form a natural framework for examination. It turns out that the concept of *Hardy's regular convergence* is precisely what is needed to serve the purpose, and details in this direction are presented in the Supplement (see also Section 1.4 below).

3. We would like to point out that (1.2) represents the first and the most natural example of an infinite dimensional law of large numbers (with $\mathcal{F}$ from the classic Glivenko-Cantelli theorem for instance). One should also observe that (1.2) can be viewed as a law of large numbers in the (non-separable) Banach space of bounded functions with the supremum norm. Moreover, it may be noted that (1.2) involves the classic law of large numbers in any separable Banach space B. It clearly follows upon recalling that $\|x\| = \sup_{f^* \in S^1} |f^*(x)|$ for all $x \in B$, where S^1 denotes

the unit ball (or a smaller subset which suffices) in the dual space B^* of B. For these reasons it is clear that (1.1) may be viewed as an *infinite dimensional Birkhoff's (von Neumann's) ergodic theorem*. More details in this direction will be presented in Section 1.3 below.

4. In Section 1.1 we review the essential aspects from ergodic theory (and physics) which additionally clarify the meaning and importance of (1.1). We hope that this material will help the reader unfamiliar with ergodic theory to enter into the field as quickly as possible. For more information on this subject we further refer to the standard monographs [14, 17, 26, 36, 44, 47, 52, 70]. In Section 1.2 we review the essential facts on law of large numbers, and display its fundamental connection with ergodic theorem. This is followed in Section 1.3 by its infinite dimensional extension in the form of the uniform law of large numbers (1.2) which originated in the papers of Glivenko [34] and Cantelli [13]. For more information on the extensive subject of empirical processes we refer to the standard monographs [18, 20, 27, 28, 54, 73, 74, 88].

5. One may observe that certain measurability problems related to (1.1) and (1.2) could (and do) appear (when the supremums are taken over uncountable sets) which is due to our general hypotheses on the family $\mathcal{F}$. Despite this drawback we will often assume measurability (implicitly) wherever needed. We emphasize that this simplification is not essential and could be supported in quite a general setting by using the theory of analytic spaces. Roughly speaking, if $\mathcal{F}$ is an analytic space and the underlying random function $(x, f) \mapsto G(x, f)$ is jointly measurable, then the map $x \mapsto \sup_{f \in \mathcal{F}} G(x, f)$ is μ-measurable (see [43] p.12-14). This is a consequence of the *projection theorem*. Moreover, in this case the *measurable selection theorem* is available, as well as the *image theorem* (see [43] p.12-14). It turns out that these three facts are sufficient to support the VC theory in Chapter 3, although for simplicity we do not provide all details. Another approach could be based upon a separability assumption which would reduce the set over which the supremum is taken to a countable set. Finally, even the most general case of an arbitrary family $\mathcal{F}$ could be successfully treated using the theory of non-measurable calculus involving the upper integral (see [1] and [62]). This is done in Chapter 2, while in Chapter 4 such details are omitted.

6. There are numerous applications of the results presented and we would like to make a few general comments on this point. The uniform ergodic theorem (1.1) treated in these lecture notes appears fundamental as an infinite dimensional extension of the Glivenko-Cantelli theorem and classic ergodic theorems (see Section 1.3). A clear way of getting a good feeling for applications is first to understand the meaning of the classical Glivenko-Cantelli theorem in the i.i.d. context, and then to replace this with the context of any dynamical system. This sort of reasoning is displayed in the Supplement where the consistency problem for statistical models is treated in detail. Another class of applications comes solely from ergodic theory. For this, it is enough to think of the meaning of Birkhoff's ergodic theorem in the context of a single dynamical system, in order to see what the meaning and applications of an infinite dimensional ergodic theorem (1.1) would be in the context of a family of such systems (in *foundations of statistical mechanics* for instance). Details in this direction are presented in Section 1.4.

7. We believe that our exposition in these lectures serves a purpose for at least two reasons. First, our approach tries to unify probability theory and ergodic theory as much as possible. (We believe that this interaction is of considerable interest in general.) Second, our approach relies in part upon the theory of Gaussian processes (Chapter 3) and harmonic analysis (Chapter 4), and in part upon the theory of martingales (Supplement).

1.1 Ergodic Theorem

Before passing to an up-to-date formulation of the ergodic theorem we shall first look back at its origin. In order to understand fully the meaning of an ergodic theorem, it is indispensable to comprehend historically the first example of *Hamiltonian dynamics*. For more information in this direction we refer the reader to the excellent exposition [47].

1. The study of ergodic theory originated in problems of statistical mechanics. In this context one considers a mechanical system with d degrees of freedom, the states of which are described by values of the Hamiltonian variables:

$$q_1, \dots, q_d \ ; \ p_1, \dots, p_d \tag{1.3}$$

which form a *phase space* X being assumed a subset of $\mathbf{R}^{2d}$. For example, a system could consist of N particles (of a gas) each of which has 3 coordinates of position and 3 coordinates of momentum, so in this case $d = 3N$. By the laws of nature, the state of a system at any given time determines its state at any other time. The equations of motion of the system are:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \quad , \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \qquad (\, 1 \le i \le d \,) \tag{1.4}$$

where $H = H(q_1, \dots, q_d \ ; \ p_1, \dots, p_d)$ is the so-called *Hamiltonian function*. It follows from (1.4) that H does not depend on time:

$$\frac{dH}{dt} \equiv 0 \ . \tag{1.5}$$

Typically, the Hamiltonian function H is the sum of the kinetic and potential energy of the system:

$$H = H(q;p) = K(p) + U(q) \tag{1.6}$$

where $p = (p_1, \dots, p_d)$ is a (generalized) momentum and $q = (q_1, \dots, q_d)$ is a (generalized) position. By the theorem on the existence and uniqueness of solutions of first order differential equations, the system (1.4) determines uniquely the state $T_t(x)$ at any time t if the initial state of the system was $x = (q;p) \in X$. Thus, when time passes by, all the phase space X is transforming into itself, and this clearly happens in one-to-one and onto way. The motion generated by equations (1.4) may be called the natural motion of the system, and this motion is clearly stationary (independent of the time shift). In other words, the identity is satisfied:

$$T_t \circ T_s = T_{t+s} \tag{1.7}$$

for all $s \, , t \in \mathbf{R}$. In this way a one-parameter flow $(T_t)_{t \in \mathbf{R}}$ on the phase space X is obtained.

The special form of the system (1.4) indicates that not every transformation of the phase space into itself can appear as its natural motion. It turns out that the natural motion has some special properties. The two most important ones (upon which the whole statistical mechanics is based) are formulated in the following two theorems.

Liouville's Theorem 1.1

The Lebesgue measure of any Lebesgue measurable set $A \subset X$ *equals the Lebesgue measure of its image* $T_t(A)$ *for all* $t \in \mathbf{R}$.

Proof. Set $A_t = T_t(A)$ for $t \in \mathbf{R}$, and let λ denote Lebesgue measure in $\mathbf{R}^{2d}$. By the natural change of variables we have:

$$\lambda(A_t) = \int_{A_t} dx_1 \dots dx_{2d} = \int_A J(t; y_1, \dots, y_{2d})\ dy_1 \dots dy_{2d}$$

where $J = J(t; y_1, \dots, y_{2d})$ is the Jacobian of $(x_1, \dots, x_{2d})$ at $(y_1, \dots, y_{2d})$. Hence:

$$\frac{d}{dt}\lambda(A_t) = \int_A \frac{\partial J}{\partial t}\ dy_1 \dots dy_{2d}\ .$$

A straightforward calculation by using (1.4) shows:

$$\frac{\partial J}{\partial t} \equiv 0\ .$$

Thus $\lambda(A_t)$ is constant for all $t \in \mathbf{R}$ and equals to $\lambda(A_0) = \lambda(A)$. The proof is complete. □

Birkhoff's Ergodic Theorem (for Hamiltonian Flow) 1.2

Let A *be an invariant subset of* X *(meaning* $T_t(A) \subset A$ *for all* $t \in \mathbf{R}$*) of finite Lebesgue measure, such that it cannot be decomposed into:*

(1.8) $$A = A_1 \cup A_2$$

where both A_1 *and* A_2 *are invariant subsets of* X *of strictly positive Lebesgue measure. Then for any phase function* f *on* X *which is integrable over* A *we have:*

(1.9) $$\lim_{\Delta \to \infty} \frac{1}{\Delta} \int_0^\Delta f\big(T_t(x)\big)\ dt = \frac{1}{\lambda(A)} \int_A f(y)\ dy$$

for almost all x *in* A *with respect to the Lebesgue measure* λ .

Proof. The reader should have no difficulty in constructing the proof of this theorem in this particular case if first using the reduction to discrete time as indicated by (1.10)+(1.11) below and then applying the general version of Birkhoff's Theorem 1.6 stated and proved below. □

2. The basic idea of statistical mechanics (Gibbs) is to abandon the deterministic study of one state (which correspond to a point in the phase space) in favour of a statistical study of a collection of states (which correspond to a subset of the phase space). Liouville's Theorem 1.1 states that the flow of natural motion of a system preserves the Lebesgue measure of any measurable subset of the phase space. Birkhoff's Theorem 1.2 states if the flow of natural motion of a system satisfies property (1.8), then the time average of a phase function equals in the limit to its space average. Both of these theorems have a clear physical meaning in the context of Gibbs' fundamental approach. The first property is named "measure-preserving", and the second is called "ergodic". We shall now turn to their formal introduction by presenting a general setting which was used in the development of modern ergodic theory. In light of the preceding example we believe that the reader will have

no difficulty in comprehending the formalism of the objects and definitions to follow.

3. Before passing to formal definitions let us record yet another remark which should clarify the presentation below. It concerns an easy reduction of the continuous time case to the discrete time case in the problems on the limiting behaviour of the time average in (1.9). Due to the semigroup property (1.7) we shall now demonstrate that the reduction mentioned in essence consists only of a change of the phase function. Namely, if in the notation of Birkhoff's Ergodic Theorem 1.2 above we introduce a new phase function as follows (see Paragraph 1 in Section 3.8 below):

$$F(x) = \int_0^1 f\big(T_t(x)\big)\, dt \tag{1.10}$$

then by (1.7) we clearly have:

$$\frac{1}{N}\int_0^N f\big(T_t(x)\big)\, dt \;=\; \frac{1}{N}\sum_{k=0}^{N-1} F\big(T^k(x)\big) \tag{1.11}$$

for all $x \in X$, where we set T to denote T_1, so that T^k equals T_k. This indicates that the asymptotic behaviour of the integral in (1.11) is the same as the one of the sum in (1.11). For this reason we shall be mainly interested in discrete time formulations of the results in the sequel, but having understood the reduction principle just explained the reader will generally have no difficulty to restate these results in the continuous parameter setting.

4. **Dynamical systems**. Let $(X, \mathcal{A}, \mu)$ be a finite measure space, where with no loss of generality we assume that $\mu(X) = 1$, and let T be a measurable map from X into itself. Such a map will be called a *transformation* of X. A transformation T is said to be *measure-preserving* if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{A}$. In this case we also say that μ is T*-invariant*, and we call $(X, \mathcal{A}, \mu, T)$ a *dynamical system.* A set $A \in \mathcal{A}$ is said to be T-invariant if $T^{-1}(A) = A$. The family $\mathcal{A}_T$ of all T-invariant sets from $\mathcal{A}$ is clearly a σ-algebra on X. A measurable map f from X into $\mathbf{R}$ is said to be T-invariant if $T(f) := f \circ T = f$. It is easily verified that f is T-invariant if and only if f is $\mathcal{A}_T$-measurable.

A measure-preserving transformation T is called *ergodic* if $\mu(A)$ equals either 0 or 1 whenever $A \in \mathcal{A}_T$. Then we also say that $\mathcal{A}_T$ is trivial. Clearly T is ergodic if and only if each T-invariant map is constant μ-a.s. It is easily seen that T is ergodic if and only if:

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu\big(T^{-k}(A) \cap B\big) - \mu(A)\mu(B) \to 0 \tag{1.12}$$

for all $A, B \in \mathcal{A}$ as $n \to \infty$.

A measure-preserving T is called *strongly mixing*, if we have:

$$\mu\big(T^{-n}(A) \cap B\big) - \mu(A)\mu(B) \to 0 \tag{1.13}$$

for all $A, B \in \mathcal{A}$ as $n \to \infty$, and T is called *weakly mixing*, if we have:

$$\frac{1}{n}\sum_{k=0}^{n-1} \big|\, \mu\big(T^{-k}(A) \cap B\big) - \mu(A)\mu(B)\, \big| \to 0 \tag{1.14}$$

for all $A, B \in \mathcal{A}$ as $n \to \infty$. It is evident that every strongly mixing transformation is weakly mixing, and that every weakly mixing transformation is ergodic.

We shall conclude our definitions by stating some useful tips on ergodicity. Suppose we are given an ergodic transformation T on a probability space $(X, \mathcal{A}, \mu)$, and let A be from $\mathcal{A}$. Then if either $A \subset T^{-1}(A)$ or $T^{-1}(A) \subset A$, we can conclude that $\mu(A)$ is either 0 or 1. In particular, if $T(A) \subset A$ then $A \subset T^{-1}(A)$ and the same conclusion holds. However, if we only know that $A \subset T(A)$, then we cannot claim that $\mu(A)$ is 0 or 1. (Take for instance $A = [0, 1/2[$ in the case of a doubling transformation T in the end of this section.)

5. The earliest result on the asymptotic behaviour of the powers of a measure-preserving transformation (when the measure space is finite) is due to H. Poincaré. Despite its simplicity this result has become famous for its physical and philosophical implications (see [70] p.34-37).

Poincaré's Recurrence Theorem 1.3

Let $(X, \mathcal{A}, \mu)$ be a finite measure space (with $\mu(X) = 1$), and let T be a measure-preserving transformation of X. Then for each $A \in \mathcal{A}$ almost all points in A are recurrent. In other words, given $A \in \mathcal{A}$ one can find $N \in \mathcal{A}$ with $N \subset A$ and $\mu(N) = 0$, such that for any $x \in A \setminus N$ there exists $n \geq 1$ for which $T^n(x) \in A$.

Proof. Given $A \in \mathcal{A}$, consider the set:

$$B = A \cap \Big(\bigcap_{k=1}^{\infty} T^{-k}(X \setminus A) \Big) . \tag{1.15}$$

Then clearly B, $T^{-1}(B)$, $T^{-2}(B)$, $\ldots$ are disjoint sets, so as T is measure-preserving and μ is finite, we must have $\mu(B) = 0$. $\square$

In this context it is instructive to note that T is ergodic if and only if the mean sojourn time for almost all points in any measurable set equals the measure of the set: For any given $A \in \mathcal{A}$, there exists $N \in \mathcal{A}$ with $N \subset A$ and $\mu(N) = 0$, such that:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_A\big(T^k(x)\big) = \mu(A) \tag{1.16}$$

for all $x \in A \setminus N$.

6. Our next aim is to restate and prove Birkhoff's Ergodic Theorem 1.2. The key step in the proof is contained in the following lemma which is formulated in a somewhat more general operator setting. This formulation nevertheless seems to be the most natural one. We should recall that the linear operator T in $L^1(\mu)$ is said to be a *contraction* in $L^1(\mu)$ if $\int |Tf| \, d\mu \leq \int |f| \, d\mu$ for all $f \in L^1(\mu)$. The linear operator T in $L^1(\mu)$ is said to be *positive* if $T(f) \geq 0$ whenever $f \geq 0$. Note that the finiteness of μ is not used in the proof of the lemma, so its statement holds for infinite measure spaces too.

Maximal Ergodic Lemma 1.4

Let T be a positive contraction in $L^1(\mu)$, and let us for given $f \in L^1(\mu)$ denote $A_n = \{ \max_{1 \leq k \leq n} S_k(f) \geq 0 \}$ where $S_k(f) = \sum_{j=0}^{k-1} T^j(f)$ for $1 \leq k \leq n$. Then we have:

$$\int_{A_n} f \, d\mu \geq 0 . \tag{1.17}$$

Proof. The main point in the proof is to establish the following inequality:

$$f \geq \max_{1 \leq k \leq n} S_k(f) - T\Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ \ . \tag{1.18}$$

Once having (1.18) we can clearly conclude by using properties of T being assumed that:

$$\begin{aligned}
\int_{A_n} f \, d\mu &\geq \int_{A_n} \max_{1 \leq k \leq n} S_k(f) \, d\mu - \int_{A_n} T\Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ d\mu \\
&= \int_{A_n} \Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ d\mu - \int_{A_n} T\Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ d\mu \\
&= \int_{X} \Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ d\mu - \int_{A_n} T\Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ d\mu \\
&\geq \int_{X} \Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ d\mu - \int_{X} T\Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ d\mu \geq 0 \ .
\end{aligned} \tag{1.19}$$

To verify the validity of (1.18) it will be enough to show that:

$$f \geq S_k(f) - T\Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ \tag{1.20}$$

for all $1 \leq k \leq n$. This inequality is evident for $k = 1$. To proceed further for $1 < k \leq n$ we shall note that due to:

$$\Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ \geq S_k(f) \tag{1.21}$$

the following inequality is satisfied:

$$f + T\Big(\max_{1 \leq k \leq n} S_k(f)\Big)^+ \geq f + TS_k(f) = S_{k+1}(f) \tag{1.22}$$

for all $1 \leq k < n$. This establishes (1.20) for all $1 \leq k \leq n$, and the proof is complete. □

When specialized to measure-preserving transformations, the Maximal Ergodic Lemma 1.4 may be refined as follows.

Maximal Ergodic Inequality 1.5

Let T *be a measure-preserving transformation of a finite measure space* $(X, \mathcal{A}, \mu)$ *(with* $\mu(X) = 1$*), and let us for given* $f \in L^1(\mu)$ *and* $\lambda > 0$ *denote* $A_{n,\lambda} = \{\max_{1 \leq k \leq n} S_k(f)/k \geq \lambda\}$ *where* $S_k(f) = \sum_{j=0}^{k-1} T^j(f)$ *for* $1 \leq k \leq n$ *. Then we have:*

$$\mu\Big\{\max_{1 \leq k \leq n} S_k(f)/k \geq \lambda\Big\} \leq \frac{1}{\lambda} \int_{A_{n,\lambda}} f \, d\mu \ . \tag{1.23}$$

Proof. Note that:

$$(1.24) \qquad A_{n,\lambda} = \Big\{ \max_{1\le k\le n} S_k(f-\lambda)/k \ge 0 \Big\} = \Big\{ \max_{1\le k\le n} S_k(f-\lambda) \ge 0 \Big\} \ .$$

Hence from (1.17) we get:

$$(1.25) \qquad \int_{A_{n,\lambda}} (f-\lambda) \ d\mu \ \ge \ 0 \ .$$

However, this is precisely (1.23), so the proof is complete. $\square$

Note that by the monotone convergence theorem one can let $n \to \infty$ in both (1.17) and (1.23) and extend these inequalities to the case when the maximum is replaced by the supremum taken over all $k \ge 1$. We shall leave precise formulations of such extensions to the reader. The inequalities obtained in such a way will be freely used below.

We need the following definition to identify the limit in the next theorem. Let $(X, \mathcal{A}, \mu)$ be a finite measure space with $\mu(X) = 1$, let $\mathcal{B} \subset \mathcal{A}$ be a σ-algebra, and let $f \in L^1(\mu)$ be given. Then $E(f \mid \mathcal{B})$ denotes the *conditional expectation of* f *given* $\mathcal{B}$. It is the element of $L^1(\mu)$ which is characterized by being $\mathcal{B}$-measurable and satisfying:

$$(1.26) \qquad \int_B E(f \mid \mathcal{B}) \ d\mu \ = \int_B f \ d\mu$$

for all $B \in \mathcal{B}$. Its existence follows easily by the Radon-Nikodým theorem (see [19]).

Birkhoff's (Pointwise) Ergodic Theorem 1.6

Let T *be a measure-preserving transformation of a finite measure space* $(X, \mathcal{A}, \mu)$ *(with* $\mu(X) = 1$ *), and let* $f \in L^1(\mu)$ *be given and fixed. Then we have:*

$$(1.27) \qquad \frac{1}{n} \sum_{k=0}^{n-1} T^k(f) \to E(f \mid \mathcal{A}_T) \quad \mu\text{-a.s.}$$

as well as in $L^1(\mu)$ *as* $n \to \infty$ *.*

Proof. Denote $S_n(f) = \sum_{k=0}^{n-1} T^k(f)$ for $n \ge 1$, and introduce:

$$(1.28) \qquad f^* = \limsup_{n\to\infty} \frac{1}{n} S_n(f) \quad \text{and} \quad f_* = \liminf_{n\to\infty} \frac{1}{n} S_n(f) \ .$$

Claim 1: f^* and f_* are T-invariant functions from X into $\bar{\mathbf{R}}$. This follows in a straightforward way upon observing that:

$$(1.29) \qquad \frac{1}{n+1} S_{n+1}(f) = \frac{1}{n+1} f + \frac{n}{n+1} \Big(\frac{1}{n} S_n(f) \Big) \circ T$$

for all $n \ge 1$, and letting $n \to \infty$.

Claim 2: $-\infty < f_* \le f^* < +\infty$ μ-a.s. To see this we shall note that:

$$(1.30) \qquad \{ f^* > \lambda \} \subset \Big\{ \sup_{k\ge 1} S_k(f)/k > \lambda \Big\}$$

being valid for all $\lambda > 0$. Hence by (1.23) from the Maximal Ergodic Inequality 1.5 we get:

$$\mu\{\, f^* > \lambda \,\} \;\leq\; \frac{1}{\lambda} \int |f| \; d\mu \tag{1.31}$$

for all $\lambda > 0$. Letting $\lambda \to \infty$ we obtain $\mu\{ f^* = +\infty \} = 0$. In exactly the same way we find $\mu\{ f_* = -\infty \} = 0$. The proof of the claim is complete.

Claim 3: $f_* = f^*$ μ-a.s. To show this we shall consider the set:

$$D = \{ f_* < \alpha < \beta < f^* \} \tag{1.32}$$

where α and β , both from $\mathbf{Q}$, are given and fixed. Clearly it is enough to show that $\mu(D) = 0$.

Since D is T-invariant (it follows by Claim 1) there is no harm in replacing the original measure space $(X, \mathcal{A}, \mu)$ with its trace on D , being denoted by $(D, \mathcal{A}_D, \mu_D)$, at which we shall apply (1.17) from the Maximal Ergodic Lemma 1.4 to the functions:

$$g = (f - \beta)\, 1_D \quad \text{and} \quad h = (\alpha - f)\, 1_D \tag{1.33}$$

respectively. (Recall that $\mathcal{A}_D = \{ A \cap D \mid A \in \mathcal{A} \}$ while $\mu_D(A \cap D) = \mu(A \cap D)$ for all $A \in \mathcal{A}$.) For this note that for every $x \in D$ we have $\sup_{k \geq 1} S_k(g)(x) = \sup_{k \geq 1} (S_k(f)(x) - k\beta) > 0$. Thus the set (from the Maximal Ergodic Lemma 1.4) being defined by:

$$\Big\{ x \in D \;\Big|\; \sup_{k \geq 1} S_k(g) \geq 0 \Big\} \tag{1.34}$$

equals D itself. Applying (1.17) from the Maximal Ergodic Lemma 1.4 we can conclude:

$$\int_D (f - \beta) \; d\mu \;\geq\; 0 \; . \tag{1.35}$$

In exactly the same way we obtain:

$$\int_D (\alpha - f) \; d\mu \;\geq\; 0 \; . \tag{1.36}$$

As this would imply $\alpha \geq \beta$, we see that $\mu(D)$ must be zero. The proof of the claim is complete.

Claim 4: The sequence $\{ S_n(f)/n \mid n \geq 1 \}$ is uniformly integrable. This clearly establishes the $L^1(\mu)$-convergence in (1.27), as well as identifies the limit.

To prove the claim we shall first note that the sequence $\{ T^k(f) \mid k \geq 0 \}$ is uniformly integrable. This clearly follows from the identity:

$$\int_{\{ |f \circ T^k| > c \}} |f \circ T^k| \; d\mu = \int_{\{|f| > c\}} |f| \; d\mu \tag{1.37}$$

upon taking the supremum over all $k \geq 0$ and then letting $c \to \infty$. The claim now follows from the following simple inequalities:

$$\sup_{n \geq 1} \int_A \Big(\frac{1}{n} S_n(f) \Big) \, d\mu \;\leq\; \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} \int_A |f \circ T^k| \, d\mu \;\leq\; \sup_{k \geq 1} \int_A |f \circ T^k| \, d\mu \tag{1.38}$$

being valid for all $A \in \mathcal{A}$. The proof of the theorem is complete. $\square$

Birkhoff's Pointwise Ergodic Theorem 1.6 in its main part remains valid for infinite measure spaces as well. The a.s. convergence in (1.27) to a T-invariant limit satisfying an analogue of (1.26) for sets of finite measure is completely preserved, while the uniform integrability of the averages in (1.27) will imply the $L^1(\mu)$-convergence. Having understood basic ideas in the preceding proof, the reader will easily complete the remaining details. Note also that if T is ergodic, then $\mathcal{A}_T$ is trivial, and the limit in (1.27) equals $E(f) = \int_X f \, d\mu$ (see Paragraph 3 in Section 1.2 below).

7. Due to the linear structure of the mapping $f \mapsto f \circ T^k$ where T is a measure-preserving transformation, there are many sensible generalizations and extensions of Birkhoff's theorem in various operator or group theoretic settings. We will make no attempt here to survey this development but instead will refer the reader to the standard textbooks in ergodic theory. Still to illustrate a method developed in such a direction we shall now present the mean ergodic theorem of von Neumann, which is a little older than Birkhoff's pointwise ergodic theorem. For this we shall recall that a linear operator T in a Hilbert space H is called a *contraction* in H , if $\|Th\| \leq \|h\|$ for all $h \in H$. Each contraction T in H is clearly a bounded linear operator, or in other words $\sup_{\|h\| \leq 1} \|Th\| < \infty$. Having a bounded linear operator T in a Hilbert space H , the *dual* T^* of T is a bounded linear operator in H which is characterized by $(Tf, g) = (f, T^*g)$ being valid for all f , $g \in H$. Having a contraction T in a Hilbert space H , it is easily verified that $Tf = f$ if and only if $T^*f = f$ for $f \in H$.

Von Neumann's (Mean) Ergodic Theorem 1.7

Let T be a contraction in a Hilbert space H , and let P be a projection from H onto the closed subspace $\{ f \in H \mid Tf = f \}$. Then we have:

$$\frac{1}{n} \sum_{k=0}^{n-1} T^k(h) \to P(h) \tag{1.39}$$

for all $h \in H$ as $n \to \infty$.

Proof. The key point of the proof is the following decomposition:

$$H = \{ f \in H \mid Tf = f \} \oplus cl\big((I-T)H\big) \tag{1.40}$$

where cl denotes the closure operation in H , and $\oplus$ denotes the direct sum operation in H .

To verify (1.40) we shall denote $F = \{ f \in H \mid Tf = f \}$ and $G = cl\big((I-T)H\big)$. Thus it is enough to show that $F^\perp = G$. Since for a linear subspace S of H we have $S^{\perp\perp} = cl(S)$, it will be enough to show that $((I-T)H)^\perp = F$. For this note that $f \in ((I-T)H)^\perp$ iff $(f, (I-T)h) = (f - T^*f, h) = 0$ for all $h \in H$, iff $T^*f = f$, iff $Tf = f$, thus proving the claim as stated.

To prove the convergence in (1.39) take now any $h \in H$. Then $h = f + g$ where $f \in F$ and $g \in G$. Hence we get:

$$\frac{1}{n} \sum_{k=0}^{n-1} T^k(h) = f + \frac{1}{n} \sum_{k=0}^{n-1} T^k(g) \tag{1.41}$$

for all $n \geq 1$. By a simple approximation (which is based on the fact that the iterates of a contraction are contractions themselves) this shows that (1.39) will follow as soon as we show that the last average in (1.41) tends to zero as $n \to \infty$ for all $g \in (I{-}T)H$. So take such $g = h - Th$ for some $h \in H$. Then by a telescoping argument we find:

$$\left\| \frac{1}{n} \sum_{k=0}^{n-1} T^k(g) \right\| = \frac{1}{n} \left\| h - T^n(h) \right\| \leq \frac{2}{n} \|h\| \ . \tag{1.42}$$

for all $n \geq 1$. Letting $n \to \infty$ we obtain:

$$\frac{1}{n} \sum_{k=0}^{n-1} T^k(g) \to 0 \tag{1.43}$$

for all $g \in G$. This completes the proof. □

We shall conclude this section by recalling a very useful extension of Birkhoff's Pointwise Ergodic Theorem 1.6 that (or some of its variants) may be of particular interest for establishing uniform ergodic theorems to which these lectures are primarily devoted.

Kingman's (Subadditive) Ergodic Theorem 1.8

Let T *be a measure-preserving transformation of a finite measure space* $(X, \mathcal{A}, \mu)$ *(with* $\mu(X) = 1$ *) , and let* $\{ g_n \mid n \geq 1 \}$ *be a* T*-subadditive sequence in* $L^1(\mu)$ *, that is:*

$$g_{n+m} \leq g_n + g_m \circ T^n \quad \mu\text{-a.s.} \tag{1.44}$$

for all $n \, , \, m \geq 1$ *. Then we have:*

$$\frac{g_n}{n} \to \inf_{m \geq 1} E\left(\frac{g_m}{m} \mid \mathcal{A}_T \right) \quad \mu\text{-a.s.} \tag{1.45}$$

as $n \to \infty$ *. If* $\inf_{m \geq 1} E(g_m/m) > -\infty$ *, then the convergence in (1.45) is in* $L^1(\mu)$ *as well.*

Proof. For a proof see e.g. [19] (p.292-299). For more general formulations of this theorem see [52] (p.35-38)+(p.146-150). We shall omit details for simplicity. □

8. To indicate the existence of some other typical (abstract) examples of measure-preserving transformations which gained significant interest in the development of ergodic theory, we shall recall that a *rotation of the unit circle* $S^1 = \{ z \in \mathbf{C} : |z| = 1 \}$ is defined by the formula:

$$T_\alpha\left(e^{2\pi i \theta}\right) = e^{2\pi i (\theta + \alpha)} \tag{1.46}$$

for $\theta \in [0,1[$, where $\alpha \in [0,1[$ is given and fixed. Equivalently T_α may be regarded as a map from $[0,1[$ into $[0,1[$ defined by $T_\alpha(x) = x + \alpha$ (mod 1) being the fractional part of $x + \alpha$ for $x \in [0,1[$. Then T_α is clearly a measure-preserving transformation for all $\alpha \in [0,1[$. Moreover T_α is ergodic if and only if α is *irrational*. For a measure-theoretic proof of this fact and its standard extensions to higher dimensions see [52] (p.12-13). For a Fourier-analysis proof with standard extensions we refer the reader to [70] (p.49-51).

Further, the *doubling* transformation $T(z) = z^2$ for $z \in S^1$, or equivalently $T(x) = 2x$ (mod 1) for $x \in [0,1[$, is ergodic. It turns out to be isomorphic to the unilateral shift θ in the countable product $(\{0,1\}^{\mathbf{N}}, \mathcal{B}(\{0,1\}^{\mathbf{N}}), \mu^{\mathbf{N}})$ where $\mu(\{0\}) = \mu(\{1\}) = 1/2$. Similarly, the so-called *baker's transformation* on $[0,1[\times[0,1[$ (the action of which looks like kneading dough) is isomorphic to the bilateral shift θ in the countable product $(\{0,1\}^{\mathbf{Z}}, \mathcal{B}(\{0,1\}^{\mathbf{Z}}), \mu^{\mathbf{Z}})$ with the same μ , thus both being ergodic. For proofs and remaining details see [36].

Yet another class of interesting examples will be presented following the proof of Theorem 1.9 below, where an important link between ergodic theory and probability theory by means of shift transformations will be displayed.

Notes: Liouville's theorem goes back to the middle of the 19th century. Boltzmann (1887) introduced the word "ergodic" in connection with statistical mechanics (see [8] p.208). Poincaré [72] derived his theorem in 1899. Birkhoff [6] proved his theorem in 1931 by using a weaker maximal inequality. Von Neumann [60] deduced his theorem in 1932 via spectral theory. Von Neumann's theorem was proved first and was known to Birkhoff. The maximal ergodic lemma was first proved for measure-preserving transformations by Hopf in 1937 (see [44]). Yosida and Kakutani [95] showed in 1939 how this lemma can be used to prove Birkhoff's ergodic theorem. The short proof of the maximal ergodic lemma as given above is due to Garsia [29]. The subadditive ergodic theorem was proved by Kingman [48] in 1968. This result plays a useful role in establishing uniform convergence results as realized by Steele [79].

1.2 Law of Large Numbers

1. In order to comprehend the concept of law of large numbers suppose we consider a random phenomenon with a quantity of interest to us. Suppose moreover that as a result of our consecutive measurements of this quantity we are given numbers x_1 , x_2 , $\dots$, x_n . From the practical and theoretical point of view it would be desirable that the limit:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} x_k =: M \tag{1.47}$$

exists in some sense, and if so, its value M could be viewed as a proper (or mean) value for the quantity being measured. Modern probability theory (founded by Kolmogorov [50]) provides a mathematical framework for such and related investigations. We shall now present some basic facts in this direction.

2. The basic object in this context is the probability space $(\Omega, \mathcal{F}, P)$. In the terminology of modern measure theory $(\Omega, \mathcal{F}, P)$ is a measure space with $P(\Omega) = 1$, thus Ω is a non-empty set, $\mathcal{F}$ is a σ-algebra of subsets of Ω , and P is a normalized measure on $\mathcal{F}$. Each $\omega \in \Omega$ may be thought of as an outcome or realization of the random phenomenon. The k-th consecutive measurement of the quantity under examination is realized as a *random variable* X_k defined on Ω and taking values in $\mathbf{R}$ for $k \geq 1$. This means $X_k : \Omega \to \mathbf{R}$ is a measurable function (with respect to $\mathcal{F}$ and the Borel σ-algebra on $\mathbf{R}$ denoted by $\mathcal{B}(\mathbf{R})$) for all $k \geq 1$. In the ideal case the measurements are taken independently each from other, so that the random variables may be assumed to satisfy:

$$P\Big(\bigcap_{k=1}^{n} \{X_k \in B_k\}\Big) = \prod_{k=1}^{n} P\{X_k \in B_k\} \tag{1.48}$$

for all $B_k \in \mathcal{B}(\mathbf{R})$ and all $1 \leq k \leq n$. For this reason if (1.48) is fulfilled then $\{X_k \mid k \geq 1\}$ is said to be a sequence of *independent* random variables. Moreover, as long as we measure the same quantity in a consecutive way (with the units of time $1, 2, \dots$) we can assume that the random variables X_1 , X_2, $\dots$ are *identically distributed*:

$$P\{ X_1 \in B \} = P\{ X_2 \in B \} = \dots \tag{1.49}$$

being valid for all $B \in \mathcal{B}(\mathbf{R})$.

3. The problem (1.47) is now more precisely formulated by asking when:

$$\frac{1}{n} \sum_{k=1}^{n} X_k \to M \tag{1.50}$$

in some sense as $n \to \infty$. Clearly, since X_k's are measurable maps defined on a measure space, the classical convergence concepts from measure theory appeal naturally. The validity of (1.50) is called a *strong law of large numbers* if the convergence in (1.50) is P-a.a. (in this case we write P-a.s. and say P-almost surely), while it is called a *weak law of large numbers* if the convergence in (1.50) is in P-measure (in this case we say in P-probability) or in $L^r(P)$ for some $0 < r < \infty$. The most desirable in (1.50) is that the limit M equals the Lebesgue-Stieltjes integral of X_1 over Ω under P :

$$E(X_1) = \int_{\Omega} X_1 \, dP \ . \tag{1.51}$$

This number is called the *expectation* of X_1 . The strong law of large numbers is completely characterized by the following fundamental theorem.

Kolmogorov's (Strong) Law of Large Numbers 1.9

Let $X = \{ X_k \mid k \geq 1 \}$ be a sequence of independent and identically distributed random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. Then we have:

$$\frac{1}{n} \sum_{k=1}^{n} X_k \to E(X_1) \quad P\text{-a.s.} \tag{1.52}$$

as $n \to \infty$ if and only if $E(X_1)$ exists (its value may be $\pm\infty$).

Proof. The validity of (1.52) under the existence and finiteness of $E(X_1)$ follows from the Birkhoff's Pointwise Ergodic Theorem 1.6 as indicated in more detail following the proof. The converse follows (as in the proof of (2.132) below) by the second *Borel-Cantelli lemma* which states $P(\limsup_{n \to \infty} A_n) = 1$ whenever A_n's are independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$. The cases of infinite expectation are easily deduced from the cases of finite expectation. $\square$

4. It is evident that there is a strong analogy between the ergodic problem from the previous section and the problem of convergence in (1.50), thus (due to the more restrictive assumption of independence made here) it shouldn't be surprising that the essential part of this theorem (if $E(X_1)$ exists then (1.52) holds) follows from Birkhoff's Pointwise Ergodic Theorem 1.6. In trying to make this reduction more explicit, we shall now establish an important link between the probabilistic setting of Kolmogorov's Law 1.9 and the ergodic setting of Birkhoff's Theorem 1.6. In Example 2.28 below (see Section 2.5.4) we shall also present a major difference between them

(the characterization of (1.52) by the existence of the integral fails in Birkhoff's case).

In order to apply Birkhoff's Theorem 1.6 in the context of Kolmogorov's Law 1.9 we should chose a natural dynamical system $(X, \mathcal{A}, \mu, T)$ where $(X, \mathcal{A}, \mu)$ is a finite measure space and $T : X \to X$ is a measure-preserving transformation. To do so, first note, that the measure space should clearly be $(\Omega, \mathcal{F}, P)$. Moreover, since our interest is in establishing P-a.s. convergence in (1.52), there is no restriction to assume that $(\Omega, \mathcal{F}, P)$ equals $(\mathbf{R^N}, \mathcal{B}(\mathbf{R^N}), P_X)$ where $P_X(B) = P\{X \in B\}$ for $B \in \mathcal{B}(\mathbf{R^N})$ with $X = (X_1, X_2, \,\dots\,)$ being the given sequence of random variables, and that X_k equals the k-th projection p_k from $\mathbf{R^N}$ into $\mathbf{R}$ defined by $p_k(x_1, x_2, \,\dots\,) = x_k$ for $k \geq 1$. The state $x = (x_1, x_2, \,\dots\,)$ of the phase space $\mathbf{R^N}$ can be thought of as a realization of the (total) measurement (performed from the definite past to the indefinite future), thus after a unit of time this state is to be transformed into $(x_2, x_3, \,\dots\,)$. (The reader could note a very interesting reasoning at this step which, roughly speaking, would say that the randomness in essence is deterministic. Philosophical speculations on this point are well-known in physics.) This indicates that the transformation T above should be the (unilateral) *shift*:

$$\theta(x_1, x_2, x_3, \,\dots\,) = (x_2, x_3, \,\dots\,) \tag{1.53}$$

for $(x_1, x_2, x_3, \,\dots\,) \in \mathbf{R^N}$. In this way we have obtained $(\mathbf{R^N}, \mathcal{B}(\mathbf{R^N}), P_X, \theta)$ as a candidate for our dynamical system.

The sequence $X = \{\, X_k \mid k \geq 1 \,\}$ is called *stationary* (*ergodic*), if θ is a measure-preserving (ergodic) transformation in $(\mathbf{R^N}, \mathcal{B}(\mathbf{R^N}), P_X)$. By (1.48) and (1.49) it is easily verified that each i.i.d. sequence $X = \{\, X_k \mid k \geq 1 \,\}$ is stationary (i.i.d. means that X_k's are independent and identically distributed). Moreover, for the σ-algebra of θ-invariant sets $\mathcal{B}_\theta(\mathbf{R^N}) = \{\, B \in \mathcal{B}(\mathbf{R^N}) \mid \theta^{-1}(B) = B \,\}$ and the tail σ-algebra $\mathcal{B}_\infty(\mathbf{R^N}) = \cap_{n=1}^\infty \sigma\{X_k \mid k \geq n\}$ we have:

$$\mathcal{B}_\theta(\mathbf{R^N}) \subset \mathcal{B}_\infty(\mathbf{R^N}) \; . \tag{1.54}$$

Recalling *Kolmogorov's 0-1 law* (see [50]) which states that in the case of a sequence of independent random variables $X = \{\, X_k \mid k \geq 1\}$ the tail σ-algebra $\mathcal{B}_\infty(\mathbf{R^N})$ is P_X-trivial ($P_X(B)$ equals either 0 or 1 for all $B \in \mathcal{B}_\infty(\mathbf{R^N})$), this shows that every i.i.d. sequence $X = \{\, X_k \mid k \geq 1 \,\}$ is ergodic as well.

Applying Birkhoff's Theorem 1.6 to the dynamical system $(\mathbf{R^N}, \mathcal{B}(\mathbf{R^N}), P_X, \theta)$ with $f = p_1$, and noting that $p_1 \circ \theta^{k-1} = X_k$ for $k \geq 1$, we get:

$$\frac{1}{n} \sum_{k=0}^{n-1} p_1(\theta^k) = \frac{1}{n} \sum_{k=1}^{n} X_k \to E\big(p_1 | \mathcal{B}_\theta(\mathbf{R^N})\big) = E(X_1) \tag{1.55}$$

P-a.s. as $n \to \infty$, provided that $E|X_1| < \infty$, where the sequence $X = \{\, X_k \mid k \geq 1 \,\}$ is as in the setting of Kolmogorov's Law 1.9. This establishes (1.52) above under the assumption that $E(X_1)$ exists and is finite.

5. There are many generalizations and extensions of Kolmogorov's law of large numbers (1.52) which can be found in the standard textbooks on this subject (see [35], [75], [80] and [71]). Perhaps the best illustration of such results is another theorem of Kolmogorov which states that (1.52) above holds if $X = \{\, X_k \mid k \geq 1 \,\}$ is a sequence of independent random variables satisfying:

$$(1.56) \qquad \sum_{n=1}^{\infty} \frac{E|X_n|^2}{n^2} < \infty \ .$$

In the literature this is sometimes also referred to as Kolmogorov's (strong) law of large numbers. Note, however, that $X_1, X_2, \ \dots$ are only assumed independent, but with finite second moment.

6. **The law of iterated logarithm**. The best refinement of the law of large numbers is known as the law of the iterated logarithm. It states that for any i.i.d. sequence of random variables $X = \{ X_k \mid k \geq 1 \}$ with $S_n = \sum_{k=1}^{n} X_k$ for $n \geq 1$, we have:

$$(1.57) \qquad \limsup_{n \to \infty} \frac{S_n}{\sqrt{2n \log \log n}} = 1 \quad P\text{-a.s.}$$

if and only if $E(X_1) = 0$ and $E(X_1^2) = 1$. Clearly (1.57) is equivalent to:

$$(1.58) \qquad \liminf_{n \to \infty} \frac{S_n}{\sqrt{2n \log \log n}} = -1 \quad P\text{-a.s.}$$

No general ergodic-theoretic result of this type is possible (see [52] p.14-15).

7. **The central limit theorem**. Yet another control of the error in the law of large numbers is provided by the central limit theorem. In its simplest form it states that for any i.i.d. sequence of random variables $X = \{ X_k \mid k \geq 1 \}$ with $E(X_1) = 0$ and $E(X_1^2) = 1$, we have:

$$(1.59) \qquad \frac{S_n}{\sqrt{n}} \xrightarrow{\sim} N(0,1)$$

as $n \to \infty$, where $S_n = \sum_{k=1}^{n} X_k$ for $n \geq 1$.

The convergence in (1.59) is *convergence in distribution* (also called *weak convergence*). It means that $P\{S_n/\sqrt{n} \leq x\} \to P\{Z_1 \leq x\}$ as $n \to \infty$ for all $x \in \mathbf{R}$, where $Z_1 \sim N(0,1)$ is a normally distributed (Gaussian) random variable with expectation 0 and variance 1 , thus its distribution function is given by:

$$(1.60) \qquad F_{Z_1}(x) = P\{ Z_1 \leq x \} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \ dt$$

for $x \in \mathbf{R}$. Note that (1.59) states:

$$(1.61) \qquad \sqrt{n} \left(\frac{S_n}{n} \right) = \frac{S_n}{\sqrt{n}} \approx N(0,1)$$

for n large. Hence we see that:

$$(1.62) \qquad \frac{S_n}{n} \approx N(0, \tfrac{1}{n})$$

for n large. Note that the density function of $Z_n \sim N(0,1/n)$ (a non-negative Borel function $t \mapsto f_{Z_n}(t)$ satisfying $F_{Z_n}(x) = P\{Z_n \leq x\} = \int_{-\infty}^{x} f_{Z_n}(t) \ dt$ for all $x \in \mathbf{R}$) is given by:

$$(1.63) \qquad f_{Z_n}(t) = \sqrt{\frac{n}{2\pi}} \ \exp\left(\frac{-(\sqrt{n}t)^2}{2} \right)$$

for $t \in \mathbf{R}$, thus it is bell-shaped (even) with the maximal value $\sqrt{n/2\pi}$ (the bell-top) attained at 0 and tending to infinity, with the bell-sides approaching 0 at points different from 0 , both when $n \to \infty$. In view of (1.62) this offers a clear meaning of the central limit theorem (1.59) in the context of the law of large numbers (1.52).

The central limit theorem for weakly dependent sequences was already established by Bernstein [5], while for completely regular sequences it was fully characterized by Volkonskii and Rozanov [92]. The problem of weak convergence (central limit theorem) relative to the uniform convergence topology (in the context of empirical processes) stays out of the scope of these lectures, and for more information on this subject with additional references we refer to [3], [20], [33], [88].

Notes: The Borel-Cantelli lemma was proved by Borel [9] in 1909. Cantelli [12] noticed in 1917 that one half holds without independence (this half is now called the first Borel-Cantelli lemma). Kolmogorov [49] showed in 1930 that (1.56) is sufficient for the strong law of large numbers if the variables are independent. Kolmogorov [50] proved his strong law 1.9 in 1933. The "if part" also follows from Birkhoff's ergodic theorem 1.6 by using Kolmogorov's 0-1 law [50] as shown above. Weak laws appeared much earlier, and go back to Bernoulli (1713). The law of iterated logarithm is in essence discovered by Khintchine in 1923 (see [45] and [46]). The "if part" for (1.57) in full generality (of an i.i.d. sequence with finite second moment) was first proved in 1941 by Hartman and Wintner [38], while the "only if" part was established in 1966 by Strassen [81]. The central limit theorem goes back to de Moivre (1733) and Laplace (1812). Lindeberg [55] first proved a theorem in 1922 which contains the central limit theorem (1.59).

1.3 Infinite Dimensional Extensions

1. The following important extension of Kolmogorov's law of large numbers (1.52) is established when considered in the particular setting of empirical distribution functions.

Glivenko-Cantelli's Theorem 1.10

Let $X = \{ X_k \mid k \geq 1 \}$ *be a sequence of independent and identically distributed random variables defined on the probability space* $(\Omega, \mathcal{F}, P)$ *. Then we have:*

$$\sup_{x \in \mathbf{R}} \left| \frac{1}{n} \sum_{k=1}^{n} 1_{]-\infty,x]}(X_k) - P\{X_1 \leq x\} \right| \to 0 \quad P\text{-}a.s. \tag{1.64}$$

as $n \to \infty$.

Proof. This result follows from Corollary 3.26 and Example 3.30 below. More straightforward proofs can also be found in standard textbooks (see e.g. [19] p.314). □

Note that for each fixed $x \in \mathbf{R}$ we have P-a.s. convergence in (1.64) by Kolmogorov's Law 1.9. The novelty in (1.64) is that this convergence is uniform over all $x \in \mathbf{R}$. For its profound implications this result is often referred to as *the Fundamental Theorem of Statistics*. Not only that the *empirical distribution* (on the left) approaches the true distribution (on the right):

$$\frac{1}{n} \sum_{k=1}^{n} 1_{]-\infty,x]}(X_k) \to P\{X_1 \leq x\} \quad P\text{-a.s.} \tag{1.65}$$

for each fixed $x \in \mathbf{R}$, but this also happens uniformly over all $x \in \mathbf{R}$ as stated in (1.64).

The importance of this conclusion for statistical inference is evident: Consecutive measurements x_1 , x_2 , ... , x_n of the quantity of interest associated with the random phenomenon (in the beginning of Section 1.2 above) can be used to determine the distribution law of this quantity as accurately (in probability) as needed. The random variable appearing in (1.64) is called *the Kolmogorov-Smirnov statistics*. For more information in this direction see [21].

2. On the other hand it was natural to ask if Kolmogorov's law of large numbers (1.52) also holds when observations $x_1 = X_1(\omega)$, $x_2 = X_2(\omega)$, ... , $x_n = X_n(\omega)$ are taken with values in a more general set being equipped with a structure which makes such a convergence meaningful (linear topological space). A first definite answer to this question was obtained by Mourier in 1951 (see [58] and [59]) in the case when X_k's take values in a *separable* Banach space B . Along these lines a lot of people entered into investigations of such and related problems. This field is called *Probability in Banach spaces*, and for a review of results established and methods developed we refer the reader to [54].

However, the assumption of separability of B turns out too restrictive even to cover the first and the most natural example of infinite dimensional law of large numbers (1.64). To see this, note that the underlying Banach space in (1.64) is $(B(\mathbf{R}), \|\cdot\|_\infty)$, where $B(\mathbf{R})$ denote the linear space of all bounded functions on $\mathbf{R}$, and the *sup-norm* is defined by $\|f\|_\infty = \sup_{x\in\mathbf{R}} |f(x)|$ for $f \in B(\mathbf{R})$. Then $\{ 1_{]-\infty,\,\cdot\,]}(X_n) \mid n \geq 1\}$ can be viewed as an i.i.d. sequence of random variables taking values in $B(\mathbf{R})$, and the P-a.s. convergence in (1.64) is exactly P-a.s. convergence in (1.52) taken to hold in the Banach space $(B(\mathbf{R}), \|\cdot\|_\infty)$. A well-known fact is that $B(\mathbf{R})$ is not separable with respect to the sup-norm $\|\cdot\|_\infty$ (take for instance $F = \{ f_x \mid x \in \mathbf{R} \}$ with $f_x = 1_{]-\infty,x]}$, then $\|f_{x'} - f_{x''}\|_\infty = 1$ for $x' \neq x''$, while F is uncountable). Thus the Mourier result cannot be applied in the context of the Glivenko-Cantelli theorem (1.64). In Chapter 2 and Chapter 3 we shall present results which do apply in this context. Such statements are commonly called *uniform laws of large numbers*.

3. Going back to ergodic theory we have seen in Section 1.1 how Birkhoff's Pointwise Ergodic Theorem 1.6 was extended to Von Neumann's Mean Ergodic Theorem 1.7 in the operator setting of Hilbert spaces. In 1938 Yosida [94] further extends this result by proving a *mean* ergodic theorem in Banach spaces. These theorems are pointwise ergodic theorems, and the convergence is taken with respect to the norm in the (Hilbert) Banach space. For a review of the results which followed, particularly in the setting of various specific spaces, we refer the reader to [52].

In 1941 Yosida and Kakutani [96] went further and obtained sufficient conditions on a bounded linear operator T in a Banach space B for the averages $(1/n)\sum_{k=0}^{n-1} T^k$ converge to an operator P in the operator norm (uniform operator topology):

$$\left\| \frac{1}{n} \sum_{k=0}^{n-1} T^k - P \right\| \to 0 \tag{1.66}$$

as $n \to \infty$. Since the convergence in (1.66) is taken with respect to the uniform operator topology, this result has been referred to as the *uniform ergodic theorem*, and has been well applied to some problems in Markov processes theory.

However, when B is taken to be $L^1(\mu)$ and T is the composition with a measure-preserving transformation, that is $T(f) = f \circ T$ for $f \in L^1(\mu)$, then T satisfies (1.66) if and only if T is periodic with bounded period. This shows that the concept of uniform ergodic theorem in the sense of Yosida and Kakutani does not apply in the context of a dynamical system. In other words, the *operator* uniform ergodic theorem (1.66) does not provide applications to the uniform law of large numbers, and in particular, cannot be applied in the context of the Glivenko-Cantelli theorem (1.64) (with or without the assumption of independence). This is the crucial point which motivated our exposition in these lectures.

4. Thus, to summarize, we would like to point out the following facts: (1) We have Birkhoff's ergodic theorem (for dynamical systems) and Kolmogorov's law of large numbers (i.i.d. finite dimensional case) as a special case of it. These two theorems are at the basis of modern ergodic and probability theory. (2) We have the Glivenko-Cantelli theorem (i.i.d. infinite dimensional case) which is the first and the most natural example of an infinite dimensional law of large numbers. This theorem is at the basis of modern statistics. (3) We have the uniform law of large numbers (i.i.d. infinite dimensional case) which contains the GC theorem as a special case. These include the best known and the most natural examples of infinite dimensional laws of large numbers. (4) We do not have a uniform ergodic theorem which contains the uniform law of large numbers as a special case (as is the case in (1) above with Birkhoff's ergodic theorem and Kolmogorov's law of large numbers). Our lectures are devoted to precisely such a class of theorems. In our opinion this is the most natural development:

Kolmogorov's law —> *Birkhoff's ergodic theorem*

GC's theorem —> *Uniform ergodic theorem* (*for dynamical systems*)

The fact is that the Yosida-Kakutani (operator norm) theorem does not cover the case of dynamical systems, and moreover does not include either uniform law of large numbers or the GC theorem, although it is widely called the uniform ergodic theorem. It should be noted however, which is seen as follows, that this theorem is indeed a uniform ergodic theorem in our terminology as well. Recalling in the setting above that $\|x\| = \sup_{f^* \in S^1} |f^*(x)|$ and $\|T\| = \sup_{\|x\| \leq 1} \|Tx\|$, we see that the mean ergodic theorem may be seen as a uniform ergodic theorem over the unit ball S^1 in the dual space B^* consisting of all continuous linear functionals on B, and the uniform ergodic theorem in the sense of Yosida and Kakutani can be viewed as a uniform ergodic theorem over the unit ball in the Banach space B. For more details and more general formulations of these problems see [67]. The essential facts and results on operators and functionals in Banach spaces used above can be found in the standard monograph on the subject [17].

Notes: The Glivenko-Cantelli theorem was first proved by Glivenko [34] in 1933 for the sample from a continuous distribution function, and then in the same year it was extended by Cantelli [13] to the general case. Tucker [86] proves that the GC theorem remains valid for (strictly) stationary sequences which are not necessarily ergodic, thus the limit being a random variable (the conditional distribution with respect to the σ-algebra of shift-invariant sets). It was an interesting question to see which classes of sets would further admit such an extension. Stute and Schuman [82] extended Tucker's result to $\mathbf{R}^k$ and showed that the empirical measures converge uniformly over intervals, half-spaces and balls. Their proof is interesting since it uses a well-known fact that each such (strictly) stationary sequence can be decomposed into ergodic components (see [76] p.171-178). For this reason we are mainly interested in ergodic sequences in the text below.

1.4 Applications to dynamical systems

In this section we shall describe some fundamental applications of the uniform ergodic theorem (1.1) to dynamical systems. There are three different types of applications we want to display, the third one being towards consistency of statistical models, the greater details of which are given in the Supplement. The first two types will be explained in the context of a dynamical system as considered in the beginning of Section 1.1.

1. Suppose we are given a physical system with states belonging to the phase space X and collections of states belonging to the σ-algebra $\mathcal{A}$ of subsets of X. By the law of nature the phase space X transforms into itself through a one-parameter flow $(T_t)_{t\in\mathbf{R}}$ as explained in (1.7). Due to the reduction principle (1.10)-(1.11), there will be no restriction to assume that the units of time are discrete. Suppose moreover that we are given a probability measure μ on $(X,\mathcal{A})$ for which $T := T_1$ is measure-preserving and ergodic. Then by Birkhoff's Theorem 1.6 we have:

$$(1.67) \qquad \frac{1}{n}\sum_{k=0}^{n-1} 1_A\big(T^k(x)\big) \rightarrow \mu(A)$$

as $n \rightarrow \infty$ for μ-a.a. $x \in X$. This shows that the measure μ can be thought of as a probability measure which describes how likely states of the system appear: The value of μ at any (measurable) collection of states may be interpreted as the appearance probability of any state from this collection. (Recall also (1.16) above.)

2. Now, if the measure μ is not known a priori, the question of how to find it appears fundamental. We may observe that (1.67) can be used to determine its value at any given and fixed $A \in \mathcal{A}$ as accurate as needed: Given ε_1, $\varepsilon_2 > 0$, there is an $n_0 \geq 1$ large enough, such that for all $n \geq n_0$ the probability of the set of all $x \in X$ for which:

$$(1.68) \qquad \left| \frac{1}{n}\sum_{k=0}^{n-1} 1_A\big(T^k(x)\big) - \mu(A) \right| \geq \varepsilon_2$$

is smaller than ε_1. A serious disadvantage of this procedure is that the n_0 found also depends on the set A from $\mathcal{A}$, so to determine the measure μ completely within a given and fixed level of accuracy, it would mean that we have to possess information on the entire collection of time averages (from the definite past to the indefinite future). This is clearly not satisfactory, so what we would really want is that (1.68) holds in the following stronger form: Given $\varepsilon_1, \varepsilon_2 > 0$, there is an $n_0 \geq 1$ large enough, such that for all $n \geq n_0$ the probability of the set of all $x \in X$ for which:

$$(1.69) \qquad \sup_{A\in\mathcal{C}} \left| \frac{1}{n}\sum_{k=0}^{n-1} 1_A\big(T^k(x)\big) - \mu(A) \right| \geq \varepsilon_2$$

is smaller than ε_1, where $\mathcal{C} \subset \mathcal{A}$ is a family of sets large enough so that a probability measure on $\mathcal{A}$ is uniquely determined by its values on $\mathcal{C}$. It is well-known that any family $\mathcal{C}$ which generates $\mathcal{A}$ (meaning that $\mathcal{A}$ is the smallest σ-algebra on X which contains $\mathcal{C}$), and which is closed under finite intersections, satisfies this property (see [19]). For example, the family $\mathcal{C} = \{\,]-\infty, x] \mid x \in \mathbf{R}\,\}$ satisfies this property relative to $\mathcal{A}$ as the Borel σ-algebra on $\mathbf{R}$.

The preceding analysis shows that the uniform ergodic theorem:

$$\sup_{A \in \mathcal{C}} \left| \frac{1}{n} \sum_{k=0}^{n-1} 1_A\big(T^k\big) - \mu(A) \right| \rightarrow 0 \quad \mu\text{-a.s.} \tag{1.70}$$

plays a fundamental role in determining the measure μ, which further plays the essential role in our mathematical description of the physical system. Clearly, such uniform ergodic theorems are at the basis of statistical inference for dynamical systems, much in the same manner as the Glivenko-Cantelli theorem is at the basis of modern statistics for i.i.d. observations.

3. Once the measure μ is known, the essential task in a mathematical description of the given physical system is to find and investigate conditions under which the time averages of a phase function can be replaced by the phase average of the same function. The importance and meaning of this task will be now explained.

With our physical system we usually associate some physical quantities of interest (energy, temperature, pressure, density, colour, etc.). The values of these quantities characterize the state of the system, and in turn are uniquely determined by this state. Thus a physical quantity appears as a phase function $f : X \rightarrow \mathbf{R}$, which is measurable with respect to $\mathcal{A}$. Therefore, if we wish to compare deductions of our theory with the experimental data from measurements of the physical quantity, we would have to compare the values of the physical quantity found experimentally with the values of the phase function f furnished by our theory. Very often, for practical reasons, the latter cannot be done (for systems in statistical mechanics, for instance, where the number of degrees of freedom is astronomical), and in such a case the time average of the phase function:

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k) \tag{1.71}$$

is seen as a natural interpretation of experimental measurements. Thus, in such a case, we will have to compare experimental data not with separate values of the phase function, but with their averages over very large intervals of time. (Collisions of molecules occur so rapidly that no matter how small interval of time our experiment has been made through, the time effect appears astronomical.) Knowing that the time averages (1.71) are close to the space average of the phase function:

$$\int_X f \ d\mu \tag{1.72}$$

we reduce our problem to comparing the experimental data with the space average (1.72), the evaluation of which is accessible to the methods of mathematical analysis. For these reasons the space average is usually taken as a theoretical interpretation of the physical quantity.

The mathematical task mentioned above can (to a certain extent) be solved by means of Birkhoff's Theorem 1.6: If T is ergodic, then the time average coincides with the space average:

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k) \approx \int_X f \ d\mu \tag{1.73}$$

for n large. This approach reduces the problem to verifying if T is ergodic, and as soon as we know that this is true, our mathematical task will be completed.

4. In some cases of interest (for instance when the system is not isolated) the physical quantity under consideration may depend on a parameter $\lambda \in \Lambda$ (which may be a priori unknown to us) through the phase function $f(\,\cdot\,,\lambda) : X \to \mathbf{R}$. In view of the exposition above, two distinct problems appear naturally in this context. The first one deals with *stability* of the system on the parameter $\lambda \in \Lambda$. The second one is the statistical problem of determining the *true* (but unknown) parameter λ_* which correspond to the physical quantity under observation. Below we describe how both of these problems require knowledge of the uniform ergodic theorem for the dynamical system $(X, \mathcal{A}, \mu, T)$ relative to the family of phase functions $\mathcal{F} = \{\, f(\,\cdot\,,\lambda) \mid \lambda \in \Lambda \,\}$:

$$\sup_{\lambda \in \Lambda} \left| \frac{1}{n} \sum_{k=0}^{n-1} f(T^k, \lambda) - \int_X f(x,\lambda)\ \mu(dx) \right| \to 0 \quad P\text{-a.s.} \tag{1.74}$$

as $n \to \infty$.

The first problem on stability of the system on the parameter $\lambda \in \Lambda$ is motivated by our desire to have the time average and the space average equal:

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k, \lambda) \ \approx \int_X f(x,\lambda)\ \mu(dx) \tag{1.75}$$

not just for each individual $\lambda \in \Lambda$, but simultaneously for all them, when n is large. The meaning of this is clear: Experimental measurements of the physical quantities do not depend on a particular quantity ($\lambda \in \Lambda$) but are at once valid for all of them. In other words, whenever the uniform ergodic theorem (1.75) is valid for a family of phase functions, the physical system is (experimentally) stable on the physical quantities which are represented by these phase functions. This fact is of great interest in foundations of statistical mechanics for instance.

The second problem is of a statistical nature and is concerned with finding the true parameter λ_* which correspond to the phase function representing the physical quantity we observe. In this problem, contrary to the first problem, the time average:

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k, \lambda) \tag{1.76}$$

must be available, thus the considerations below are not suitable for systems with a large number of degrees of freedom for instance. The essential idea of the method is based upon the fact that the true parameter λ_* may be characterized in terms of the space average:

$$\int_X f(x,\lambda)\ \mu(dx) \tag{1.77}$$

being considered as a function of $\lambda \in \Lambda$. Applying the same characterization to the time average (1.76), we may obtain a good approximation for the true parameter value. For instance, the true parameter λ_* is in some cases a maximum point of the space average (1.77) on Λ , and the validity of the uniform ergodic theorem (1.74) will then guarantee that the maximum points $\hat{\lambda}_n$ of the time average in (1.76) approach the maximum point of the limit (1.77), thus the true parameter value λ_* . This idea is illustrated in greater detail through a specific example in the Supplement.

5. Having understood the basic ideas and principles presented above, we hope that the reader will be able to recognize further applications of the uniform ergodic theorem for dynamical systems in a different context as well.

Notes: For more details and facts which will additionally clarify and deepen the meaning of the applications presented above, the reader is referred to the fundamental book of Khintchine [47]. Kryloff and Bogoliouboff [53] develop a theory where they start with a measure-preserving transformation T and try to define an ergodic T-invariant measure μ by (1.67) when the limit exists. Fisher [24]-[25] presents foundations of theoretical statistics and theory of statistical estimation which are closely related to the statistical applications of the uniform ergodic theorem presented above.

2. Metric Entropy With Bracketing Approach

In this chapter we present a metric entropy with bracketing approach towards uniform laws of large numbers and uniform ergodic theorems for dynamical systems. The first section concerns the Blum-DeHardt law of large numbers which appeared in [7] and [15]. The second section is devoted to its extension which is due to Hoffmann-Jørgensen [39]. Our exposition in this context follows the stationary ergodic case as appeared in [66]. We return to the i.i.d. case in the fifth section. The third section presents a uniform ergodic lemma which was proved by the author in [67]. The fourth section relies upon this result and presents a uniform ergodic theorem for dynamical systems, taken also from [67], which extends the results from the first two sections to the case of general dynamical systems. Although there are other ways to prove the essential part of this theorem (as indicated below), we think that the approach relying upon the uniform ergodic lemma is the most natural one (recall that the proof of classic Birkhoff's Ergodic Theorem 1.6 relies upon the Maximal Ergodic Lemma 1.4, and note that the uniform ergodic lemma is designed to play a similar role in the case of uniform convergence). The fifth section is reserved for examples and complements.

The following definition plays a key role throughout Chapter 2 and Chapter 3. Its concept should always be compared with related concepts whenever appear.

Definition (Metric Entropy)

Suppose we are given a pseudo-metric space (T,d) *and a subset* A *of* T *, and let* $\varepsilon > 0$ *be given and fixed. Then the covering number* $N(\varepsilon, A; d)$ *is defined as the smallest number* $N \geq 1$ *such that for some* $t_1, \ldots, t_N$ *in* A *we have* $\min_{1\leq k\leq N} d(t,t_k) \leq \varepsilon$ *for all* $t \in A$ *. The number* $\log N(\varepsilon, A; d)$ *is called the metric entropy of* A *with respect to* d *.*

The concept of metric entropy was introduced by Kolmogorov and Tikhomirov in 1959 (see [51]). It has proved a useful measure for the size of subsets in pseudo-metric spaces.

2.1 The Blum-DeHardt law of large numbers

Let $(S, \mathcal{A}, \pi)$ be a probability space, and let $\mathcal{F} \subset L^1(\pi)$ be a family of functions. Given $g \leq h$ in $L^1(\pi)$, we denote $[g,h] = \{\, f \in L^1(\pi) \mid g \leq f \leq h \,\}$. Given $\varepsilon > 0$, we let $N_1[\varepsilon, \mathcal{F}]$ denote the smallest $N \geq 1$ such that for some $g_1 \leq h_1, \ldots, g_N \leq h_N$ in $L^1(\pi)$ we have:

$$\mathcal{F} \subset \textstyle\bigcup_{k=1}^{N} [g_k, h_k] \tag{2.1}$$

$$\max_{1\leq k\leq N} \int_S (h_k - g_k)\; d\pi \;<\; \varepsilon\ . \tag{2.2}$$

The set $[f,g]$ is called a *bracket*, and $\log N_1[\,\cdot\,, \mathcal{F}]$ is called a *metric entropy with bracketing* (according to Dudley [18]). The main result of this section is stated as follows.

Theorem 2.1 (Blum 1955, DeHardt 1971)

Let $\{\xi_j\}_{j\geq 1}$ *be a sequence of independent and identically distributed random variables defined on a probability space* $(\Omega, \mathcal{F}, P)$ *with values in the measurable space* $(S, \mathcal{A})$ *and distribution law* π *. If* $\mathcal{F} \subset L^1(\pi)$ *satisfies* $N_1[\varepsilon, \mathcal{F}] < \infty$ *for all* $\varepsilon > 0$ *, then we have:*

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j) - Ef(\xi_1) \right| \to 0 \quad P\text{-a.s.}$$

as $n \to \infty$ *.*

Proof. Given $\varepsilon > 0$, put $N = N_1[\varepsilon, \mathcal{F}]$. Then there exist $g_1 \le h_1, \ldots, g_N \le h_N$ in $L^1(\pi)$ satisfying (2.1) and (2.2). Given $f \in \mathcal{F}$, by (2.1) there exists $1 \le k \le N$ such that $g_k \le f \le h_k$. Moreover, we have:

$$
\begin{aligned}
(2.3) \qquad \frac{1}{n}\sum_{j=1}^{n}\big(f(\xi_j) - Ef(\xi_1)\big) &= \frac{1}{n}\sum_{j=1}^{n}\big(f(\xi_j) - g_k(\xi_j)\big) + \frac{1}{n}\sum_{j=1}^{n}\big(g_k(\xi_j) - Eg_k(\xi_1)\big) \\
&\quad + Eg_k(\xi_1) - Ef(\xi_1)
\end{aligned}
$$

for all $n \ge 1$. From (2.3) we obtain:

$$
\begin{aligned}
(2.4) \qquad \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{j=1}^{n} f(\xi_j) - Ef(\xi_1) \right| &\le \max_{1\le k\le N} \frac{1}{n}\sum_{j=1}^{n}\big(h_k(\xi_j) - g_k(\xi_j)\big) \\
&+ \max_{1\le k\le N} \left| \frac{1}{n}\sum_{j=1}^{n} g_k(\xi_j) - Eg_k(\xi_1) \right| + \max_{1\le k\le N} E\big(h_k(\xi_1) - g_k(\xi_1)\big)
\end{aligned}
$$

for all $n \ge 1$. From (2.2) and (2.4) we see by Kolmogorov's Law 1.9 that:

$$
\limsup_{n\to\infty} \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{j=1}^{n} f(\xi_j) - Ef(\xi_1) \right| \le 2 \max_{1\le k\le N} \int_S (h_k - g_k)\ d\pi < 2\varepsilon \quad P\text{-a.s.}
$$

The proof is completed by letting $\varepsilon \downarrow 0$ over rationals. $\square$

The next two interesting facts are from [18]. The first shows that the simple sufficient condition from Theorem 2.1 may provide applications of considerable interest.

Proposition 2.2
Let B be a separable Banach space with the norm $\|\cdot\|$, and let π be a probability measure on the Borel σ-algebra of B satisfying:

$$
(2.5) \qquad \int_B \|x\|\ \pi(dx) < \infty .
$$

If $\mathcal{F}$ is the unit ball in the dual space B^* of B , then $N_1[\varepsilon, \mathcal{F}] < \infty$ for all $\varepsilon > 0$.

Proof. Let $\varepsilon > 0$ be given and fixed. Since π is tight (see [19]), then by (2.5) there exists a compact set $K = K_\varepsilon$ in B such that $\int_{B\setminus K} \|x\|\,\pi(dx) < \varepsilon/6$. Consider the family $\mathcal{G}$ consisting of all f's restricted to K when f runs over $\mathcal{F}$. Then $|g(x)| \le \|x\| \le M < \infty$ for all $g \in \mathcal{G}$ and all $x \in K$. Hence we see that the family $\mathcal{G}$ is equicontinuous and uniformly bounded with respect to the supremum norm $\|\cdot\|_K$ on K . Therefore by Arzelà-Ascoli's theorem (see [19]) the family $\mathcal{G}$ is totally bounded. Thus there exist $f_1, \ldots, f_N \in \mathcal{F}$ such that:

$$
\sup_{f\in\mathcal{F}} \inf_{1\le k\le N} \|f - f_k\|_K < \varepsilon/3 .
$$

Now, define $g_k(x) = (f_k(x) - \varepsilon/3)\cdot 1_K(x) - \|x\|\cdot 1_{B\setminus K}(x)$ and $h_k(x) = (f_k(x) + \varepsilon/3)\cdot 1_K(x) + \|x\|\cdot 1_{B\setminus K}(x)$ for all $x \in B$ and all $k = 1, \ldots, N$. Then for any $f \in \mathcal{F}$ with $\|f - f_k\|_K < \varepsilon/3$ for some $1 \le k \le N$, we have $g_k(x) \le f(x) \le h_k(x)$ for all $x \in B$. Moreover, we have:

$$\int_B \big(h_k(x) - g_k(x)\big)\ \pi(dx) = \int_{B\setminus K} 2\,\|x\|\ \pi(dx) + \int_K \frac{2\varepsilon}{3}\ \pi(dx) \leq \varepsilon$$

for all $1 \leq k \leq N$. Thus $N_1[\varepsilon, \mathcal{F}] \leq N < \infty$, and the proof is complete. □

From the preceding two results we obtain Mourier's classic law of large numbers [58]-[59] (see Section 1.3). The results of this section do also extend to more general U-statistics (see [2]).

Corollary 2.3 (Mourier 1951)
Let $\{X_j\}_{j\geq 1}$ be a sequence of independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in the separable Banach space B with the norm $\|\cdot\|$. If the following condition is satisfied:

$$\int_\Omega \|X\|\ dP < \infty \tag{2.6}$$

then there exists $EX_1 \in B$ such that:

$$\frac{1}{n}\sum_{j=1}^{n} X_j \to EX_1 \quad P\text{-a.s.} \tag{2.7}$$

as $n \to \infty$.

Proof. Put $S_n = \sum_{j=1}^{n} X_j$ for $n \geq 1$. Then by Theorem 2.1 and Proposition 2.2 we see from (2.6) that the sequence S_n/n is P-a.s. Cauchy's sequence in B . Thus $S_n/n \to Z$ P-a.s., where Z is a random variable from Ω into B . Hence by Kolmogorov's Law 1.9 we easily get $Ef(X_1) = f(Z)$ P-a.s. for all $f \in \mathcal{F}$. (Here $\mathcal{F}$ denotes the unit ball in the dual space B^* of B .) Thus $Z \in f^{-1}\big(\{Ef(X_1)\}\big)$ P-a.s. for all $f \in \mathcal{F}$.

Let now $\mathcal{F}_0$ be a countable subset of $\mathcal{F}$ satisfying $\|x\| = \sup_{f\in\mathcal{F}_0} |f(x)|$ for all $x \in B$. Such a set exists by the Hahn-Banach theorem (see [17]). Let $G = \bigcap_{f\in\mathcal{F}_0} f^{-1}\big(\{Ef(X_1)\}\big)$. Then $Z \in G$ P-a.s., and moreover G consists of a single point. For the second claim note that $x, y \in G$ implies $f(x) = f(y)$ for all $f \in \mathcal{F}_0$, and thus $\|x-y\| = \sup_{f\in\mathcal{F}_0} |f(x-y)| = 0$. Thus (2.7) is satisfied, and the proof is complete. □

2.2 Extension to necessary and sufficient conditions

Our main aim in this section is to present a necessary and sufficient condition for the uniform law of large numbers (called *eventually totally bounded in the mean*) which includes the Blum-DeHardt sufficient condition from Theorem 2.1 as a particular case. The result in the i.i.d. case is due to Hoffmann-Jørgensen [39]. Our exposition follows the stationary ergodic case as appeared in [66]. Further necessary and sufficient conditions are presented as well.

1. **Stationary ergodic sequences.** We begin by considering stationary sequences of random variables. Roughly speaking, these sequences may be described as those whose distributions remain unchanged as time passes by (see Section 1.2). More precisely, a sequence of random variables $\{\ \xi_j \mid j \geq 1\ \}$ is said to be *stationary*, if we have:

$$(\xi_{n_1}, \ldots, \xi_{n_k}) \sim (\xi_{n_1+p}, \ldots, \xi_{n_k+p}) \tag{2.8}$$

for all $1 \le n_1 < \ldots < n_k$ and all $p = 1, 2 \ldots$. In order to explain the concept of a stationary sequence in more detail, we shall recall some concepts from ergodic theory (see Section 1.1).

Let $(\Omega, \mathcal{F}, P)$ be a probability space, then a map Θ from Ω into Ω is said to be a *measure-preserving* transformation, if it is measurable and satisfies $P \circ \Theta^{-1} = P$. Measure-preserving transformations are also called *endomorphisms*. Let Θ be an endomorphism in $(\Omega, \mathcal{F}, P)$, then a set A in $\mathcal{F}$ is said to be Θ*-invariant*, if $\Theta^{-1}(A) = A$. The family $\mathcal{F}_\Theta$ of all Θ-invariant sets is a σ-algebra in Ω . An endomorphism Θ is called *ergodic*, if $P(A) \in \{0, 1\}$ for all $A \in \mathcal{F}_\Theta$. A measurable function f from Ω into $\mathbf{R}$ is called Θ*-invariant*, if $f \circ \Theta = f$. It is easily verified that f is Θ-invariant, if and only if f is $\mathcal{F}_\Theta$-measurable. Therefore, if Θ is ergodic, then any Θ-invariant function is equal to a constant P-a.s. An endomorphism Θ in $(\Omega, \mathcal{F}, P)$ is called *strongly mixing*, if we have:

$$\lim_{n \to \infty} P(A \cap \Theta^{-n}(B)) = P(A) \cdot P(B)$$

for all $A, B \in \mathcal{F}$. Clearly, if Θ is strongly mixing, then it is ergodic as well.

Let us now consider a measurable space $(S, \mathcal{A})$, and let $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$ denote the countable product of $(S, \mathcal{A})$ with itself. Then *the unilateral shift* θ is a map from $S^{\mathbf{N}}$ into $S^{\mathbf{N}}$ defined by:

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

for all $(s_1, s_2, \ldots) \in S^{\mathbf{N}}$. Let π be a probability measure on $(S, \mathcal{A})$, and let $\pi^{\mathbf{N}}$ be the countable product of π with itself. Then θ is an endomorphism in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}}, \pi^{\mathbf{N}})$, and it is well-known that θ is strongly mixing and thus ergodic (see Section 1.2).

2. We are now in position to define stationarity in an instructive way. Let $\xi = \{ \xi_j \mid j \ge 1 \}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$, and let P_ξ be the distribution law of ξ as a random variable from Ω into $S^{\mathbf{N}}$, that is $P_\xi(A) = P\{\xi \in A\}$ for all $A \in \mathcal{A}^{\mathbf{N}}$. Then P_ξ is a probability measure on $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$, and ξ is said to be *stationary*, if the unilateral shift θ is a measure-preserving transformation in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}}, P_\xi)$. It is clear that the present definition coincides with that given by (2.8). Finally, the stationary sequence ξ is called *ergodic*, if the unilateral shift θ is ergodic in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}}, P_\xi)$. The following four facts on stationary and ergodic sequences are very useful.

(2.9) **(Endomorphisms generate plenty of stationary sequences)**
Let Θ be an endomorphism in a probability space $(\Omega, \mathcal{F}, P)$, let $(S, \mathcal{A})$ be a measurable space, and let f be a measurable map from Ω into S . If we define $\xi_{i+1} = f \circ \Theta^i$ for $i \ge 0$, then $\xi = \{ \xi_j \mid j \ge 1 \}$ is a stationary sequence of random variables from Ω into S . Moreover if Θ is ergodic, then ξ is also ergodic.

(2.10) **(Shifts preserve stationarity)**
Let $\xi = \{ \xi_j \mid j \ge 1 \}$ be a stationary sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$, let θ be the unilateral shift in $S^{\mathbf{N}}$, let $(T, \mathcal{B})$ be another measurable space, and let $F : S^{\mathbf{N}} \to T$ be a measurable map. If we define:

$$\eta_{i+1} = F \circ \theta^i \circ \xi$$

for $i \geq 0$ *, then* $\eta = \{ \eta_j \mid j \geq 1 \}$ *is a stationary sequence of random variables from* Ω *into* T *. In particular, if* $\xi = \{ \xi_j \mid j \geq 1 \}$ *is a stationary sequence of real valued random variables, and if we consider:*

$$\sigma_i = \frac{1}{N} \sum_{j=i}^{i+N-1} \xi_j$$

for given and fixed $N \geq 1$ *and all* $i \geq 1$ *, then* $\sigma = \{ \sigma_j \mid j \geq 1 \}$ *is stationary.*

(2.11) **(Shifts preserve ergodicity)**
Under the assumptions in (2.10) suppose moreover that the sequence $\xi = \{ \xi_j \mid j \geq 1 \}$ *is ergodic, then* $\eta = \{ \eta_j \mid j \geq 1 \}$ *is also ergodic.*

(2.12) **(The law of large numbers for stationary sequences)**
Let $\xi = \{ \xi_j \mid j \geq 1 \}$ *be a stationary sequence of real valued random variables defined on a probability space* $(\Omega, \mathcal{F}, P)$ *, let* θ *be the unilateral shift in* $\mathbf{R}^{\mathbf{N}}$ *, and let* $\mathcal{B}_\theta(\mathbf{R}^{\mathbf{N}})$ *denote the* σ*-algebra of all* θ*-invariant sets in the Borel* σ*-algebra* $\mathcal{B}(\mathbf{R}^{\mathbf{N}})$ *. If* ξ_1 *belongs to* $L^1(P)$ *, then we have:*

$$\frac{1}{n} \sum_{j=1}^{n} \xi_j \ \rightarrow \ E\{ \xi_1 \mid \xi^{-1}(\mathcal{B}_\theta(\mathbf{R}^{\mathbf{N}}))\} \quad P\text{-a.s.}$$

as well as in P*-mean as* $n \rightarrow \infty$ *.*

Proofs of (2.9)-(2.11) are easily deduced by definition, and (2.12) follows from Birkhoff's Ergodic Theorem 1.6. We find it useful to point out that the convergence in mean is obtained from the pointwise convergence by using the following trivial but useful result on the uniform integrability of averages (see Claim 4 in the proof of Theorem 1.6):

(2.13) *Let* $\xi = \{ \xi_j \mid j \geq 1 \}$ *be a stationary sequence of real valued random variables defined on a probability space* $(\Omega, \mathcal{F}, P)$ *such that* ξ_1 *belongs to* $L^1(P)$ *. Then the sequence of averages* $\left\{ \frac{1}{n} \sum_{j=1}^{n} \xi_j \ \mid \ n \geq 1 \right\}$ *is uniformly integrable.*

3. We will now consider ergodicity of stationary sequences more closely. First, it should be noted if the sequence $\xi = \{ \xi_j \mid j \geq 1 \}$ in (2.12) is ergodic, then $\frac{1}{n} \sum_{j=1}^{n} \xi_j \rightarrow E(\xi_1)$ as $n \rightarrow \infty$. Second, we present necessary and sufficient conditions for ergodicity. Let $\xi = \{ \xi_j \mid j \geq 1 \}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$, let P_ξ denote the distribution law of ξ in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$, and let θ denote the unilateral shift in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$.

(2.14) *Suppose that* ξ *is a stationary sequence. Then* ξ *is ergodic, if and only if any of the following four equivalent conditions is satisfied:*

(i) $P_\xi(A) > 0$ *for* $A \in \mathcal{A}^{\mathbf{N}} \ \Rightarrow \ P_\xi(\bigcup_{i=0}^{\infty} \theta^{-i}(A)) = 1$

(ii) $P_\xi(A) \cdot P_\xi(B) > 0$ *for* $A, B \in \mathcal{A}^{\mathbf{N}} \ \Rightarrow \ \sum_{i=0}^{\infty} P_\xi(A \cap \theta^{-i}(B)) > 0$

(iii) $n^{-1} \sum_{j=0}^{n-1} P_\xi(A \cap \theta^{-j}(B)) \rightarrow P_\xi(A) \cdot P_\xi(B)$, *for all* $A, B \in \mathcal{A}^{\mathbf{N}}$

(iv) $n^{-1} \sum_{j=0}^{n-1} \int_{S^{\mathbf{N}}} F \cdot (G \circ \theta^j) \, dP_\xi \rightarrow \int_{S^{\mathbf{N}}} F \, dP_\xi \cdot \int_{S^{\mathbf{N}}} G \, dP_\xi$, *for all* $F, G \in L^2(P_\xi)$.

Third, we describe a connection with independence by giving a sufficient condition for ergodicity. Let p_i denote the i-th projection from $S^{\mathbf{N}}$ into S for $i \geq 1$. Then *the tail σ-algebra* in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$ is defined by $\mathcal{A}_\infty^{\mathbf{N}} = \bigcap_{n=1}^\infty \sigma(p_j \,|\, j \geq n)$, and we obviously have:

$$\mathcal{A}_\theta^{\mathbf{N}} \subset \mathcal{A}_\infty^{\mathbf{N}} \subset \mathcal{A}^{\mathbf{N}}$$

where $\mathcal{A}_\theta^{\mathbf{N}}$ denotes the σ-algebra of all θ-invariant sets in $\mathcal{A}^{\mathbf{N}}$. Therefore a triviality of the tail σ-algebra $\mathcal{A}_\infty^{\mathbf{N}}$ implies the ergodicity of the unilateral shift in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}}, P_\xi)$. It is well-known that the tail σ-algebra $\mathcal{A}_\infty^{\mathbf{N}}$ is trivial, if and only if the *Blackwell-Freedmann asymptotic independence condition* is satisfied (see [52] p.28):

$$\lim_{n\to\infty} \sup_{B \in \sigma(p_j \,|\, j \geq n)} \Big| \, P_\xi(A \cap B) - P_\xi(A) \cdot P_\xi(B) \, \Big| = 0 \tag{2.15}$$

for every $A \in \mathcal{A}_\infty^{\mathbf{N}}$. Finally, one should note that random variables from a stationary sequence are identically distributed, as well as that every sequence of *independent and identically distributed* random variables is stationary and ergodic.

4. **Non-measurable calculus.** We proceed by recalling some facts from the *calculus of non-measurable sets and functions*. For more details we refer to [1] and [62].

Let $(\Omega, \mathcal{F}, P)$ be a probability space, then P^* and P_* denote the outer and the inner P-measure. Any map Z from Ω into $\bar{\mathbf{R}}$ is called a *random element*. Then $\int^* Z \, dP$ and $\int_* Z \, dP$ denote the upper and the lower P-integral of Z , and Z^* and Z_* denote the upper and the lower P-envelope of Z (see Paragraph 2 in Section 2.3 below). Let $\{ \, Z_n \mid n \geq 1 \, \}$ be a sequence of random elements on $(\Omega, \mathcal{F}, P)$. Then the following convergence concepts show useful below:

$$Z_n \to 0 \ \ (a.s.) \text{ , if } \exists N \in \mathcal{F} \text{ such that } P(N) = 0 \text{ and } Z_n(\omega) \to 0 \text{ , } \forall \omega \in \Omega \setminus N \tag{2.16}$$

$$Z_n \to 0 \ \ (a.s.)^* \ \text{, if } \ |Z_n|^* \to 0 \ \ (a.s.) \tag{2.17}$$

$$Z_n \to 0 \ \ (P^*) \ \text{, if } \ P^*\{ \, |Z_n| \geq \varepsilon \, \} \to 0 \ , \ \forall \varepsilon > 0 \tag{2.18}$$

$$Z_n \to 0 \ \ (P_*) \ \text{, if } \ P_*\{ \, |Z_n| \geq \varepsilon \, \} \to 0 \ , \ \forall \varepsilon > 0 \tag{2.19}$$

$$Z_n \to 0 \ \ (L^1)^* \ \text{, if } \ \int^* |Z_n| \, dP \to 0 \tag{2.20}$$

$$Z_n \to 0 \ \ (L^1)_* \ \text{, if } \ \int_* |Z_n| \, dP \to 0 \ . \tag{2.21}$$

It should be noted that if Z_n is measurable for all $n \geq 1$, then the convergence concept in (2.16)-(2.17) coincides with the concept of *P-almost sure convergence*, the convergence concept in (2.18)-(2.19) coincides with the concept of *convergence in P-probability*, and the convergence concept in (2.20)-(2.21) coincides with the concept of *convergence in P-mean*. Moreover, it is easily verified that we have:

$$\begin{array}{ccc} (a.s.)^* & \Rightarrow & (a.s.) \\ \Downarrow & & \Downarrow \\ (P^*) & \Rightarrow & (P_*) \\ \Uparrow & & \Uparrow \\ (L^1)^* & \Rightarrow & (L^1)_* \end{array} \tag{2.22}$$

and no other implication holds in general.

Let ξ be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$. Then ξ is said to be *P-perfect*, if $\forall F \in \mathcal{F}$, $\exists A \in \mathcal{A}$, $A \subset \xi(F)$ such that $P(F \setminus \xi^{-1}(A)) = 0$. If ξ is P-perfect, then for any function G from S into $\bar{\mathbf{R}}$ we have $\int^* G \circ \xi \, dP = \int^* G \, dP_\xi$ and $(G \circ \xi)^* = G^* \circ \xi$. If $(\Omega, \mathcal{F}, P)$ is a probability space, then $L^{<1>}(P)$ denotes the set of all functions Z from Ω into $\mathbf{R}$ satisfying $\|Z\|_1^* = \int^* |Z| \, dP < \infty$. The space $(L^{<1>}(P), \|\cdot\|_1^*)$ is a Banach space. For more details about these facts see [62].

5. **Formulation of the problem.** First we fix some notation. If T is a non-empty set, then $\mathbf{R}^T$ denotes the set of all real valued functions defined on T, and $B(T)$ denotes the set of all *bounded* functions in $\mathbf{R}^T$. For $f \in \mathbf{R}^T$ and $A \subset T$ we put $\|f\|_A = \sup_{t \in A} |f(t)|$. Then $(f, g) \mapsto \|f - g\|_T$ defines a metric on $\mathbf{R}^T$. It is well-known that $(B(T), \|\cdot\|_T)$ is a Banach space. The finite *covering* of a set T is any family $\gamma = \{D_1, \dots, D_n\}$ of non-empty subsets of T satisfying $T = \bigcup_{j=1}^n D_j$. The family of all finite coverings of a set T will be denoted by $\Gamma(T)$.

We will now formulate the main problem under consideration and discuss some necessary conditions. Let $\xi = \{ \xi_j \mid j \geq 1 \}$ be a stationary sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$ and a common distribution law π, let T be an arbitrary set, and let $f : S \times T \to \mathbf{R}$ be a given map. Let us denote:

$$S_n(f) = \sum_{j=1}^n f(\xi_j) \tag{2.23}$$

for all $n \geq 1$. Then $S_n(f)$ maps Ω into $\mathbf{R}^T$ for $n \geq 1$, and we will study the uniform convergence of the sequence of averages $\{ \frac{1}{n} S_n(f) \mid n \geq 1 \}$ over the set T. In other words, we will look for necessary and sufficient conditions under which the uniform convergence is valid:

$$\Big\| \frac{1}{n} S_n(f) - L \Big\|_T \to 0 \quad (c) \tag{2.24}$$

for some L in $\mathbf{R}^T$ as $n \to \infty$, where (c) denotes any of the following convergence concepts: P-almost sure convergence, convergence in P-mean, convergence in P-probability. We will also consider a weak convergence concept closely related.

One may note that the map on the left-hand side in (2.24) needs not to be P-measurable under our general hypotheses on the set T. Thus we need to use concepts and results from the calculus of non-measurable sets and functions. According to (2.22) we may clearly conclude that the $(a.s.)^*$-convergence, the $(L^1)^*$-convergence and the (P^*)-convergence are convergence concepts of vital interest for (2.24). Therefore we turn our attention in this direction. All of the facts needed for establishing a weak convergence concept related to (2.24) will be presented later on. For those who would like to avoid measurability problems in (2.24) we suggest to forget all “stars” in the notation and to assume measurability wherever needed. This approach may be justified in quite a general setting by using the projection theorem as explained in Paragraph 5 of Introduction.

For the reasons stated below, the following two conditions on the map f appear natural:

$$\int^* \| f(s) \|_T \ \pi(ds) < \infty \tag{2.25}$$

$$s \mapsto f(s,t) \ \text{ is } \ \pi\text{-measurable for every } \ t \in T \, . \tag{2.26}$$

Under the canonical representation (see Section 1.2) in the case of independent and identically distributed random variables it turns out that (2.25) is necessary for the $(a.s.)$-convergence in (2.24). (This result is presented in Section 2.5.2 below.) Moreover, in this case ξ_1 is P-perfect and thus (2.25) is equivalent to the following condition:

$$\int^* \| f(\xi_1) \|_T \; dP < \infty \ . \tag{2.27}$$

However, an example due to Gerstenhaber (see Section 2.5.4 below) shows that this condition may fail in the general stationary ergodic case without additional hypothesis. In this way we are put into position to *assume that our map f satisfies (2.25)*. Of course, this is a very weak assumption and the establishment of this fact from (2.24) would have mainly a theoretical importance. Note that the condition (2.25) may be simply expressed by requiring that $\| f \|_T$ belongs to $L^{<1>}(\pi)$, as well as the condition (2.27) by requiring that $\|f(\xi_1)\|_T$ belongs to $L^{<1>}(P)$. Note also that (2.27) is a consequence of (2.25), and is equivalent to (2.25) whenever ξ_1 is P-perfect.

Let us now turn to the condition (2.26). Similarly to (2.25), under the canonical representation in the case of independent and identically distributed random variables, it turns out that (2.26) is necessary for the (P^*)-convergence in (2.24). (This result is presented in Section 2.5.3 below.) Therefore it is not too restrictive to *assume that our map f satisfies (2.26)*, and we will permanently follow this course in the sequel.

Finally, we would like to point out that *our stationary sequence ξ is supposed throughout to be ergodic*. In this case by (2.10) and (2.11) we may also conclude that the sequence $\{ f(\xi_j, t) \mid j \geq 1\}$ is stationary and ergodic for all $t \in T$. Therefore by (2.12) we get:

$$\frac{1}{n} \sum_{j=1}^{n} f(\xi_j, t) \to M(t) \quad P\text{-}a.s. \tag{2.28}$$

as $n \to \infty$, where $M(t) = \int_S f(s,t) \; \pi(ds)$ is the π-mean function of f for every $t \in T$. Thus *we may (and do) assume that L in (2.24) equals M* . Note that *(2.25) and (2.26) imply that $f(t)$ belongs to $L^1(\pi)$ for every $t \in T$* . Moreover, it is easily verified that *under (2.25) the function M belongs to $B(T)$* . This concludes our discussion on the basic hypotheses. Their necessity for (2.24) under additional hypotheses will not be considered here.

6. **The main results.** We begin by recalling that the given map f is said to be *eventually totally bounded in π-mean*, if the following condition is satisfied:

(2.29) For each $\varepsilon > 0$ there exists $\gamma_\varepsilon \in \Gamma(T)$ such that:

$$\inf_{n \geq 1} \; \frac{1}{n} \int^* \sup_{t', t'' \in A} \big| \, S_n\big(f(t') - f(t'')\big) \big| \; dP \; < \; \varepsilon$$

for all $A \in \gamma_\varepsilon$.

To be more precise, conditions (2.25) and (2.26) should also be included into definition, but for our purposes the present definition is more convenient and we shall refer to (2.25) and (2.26) separately whenever needed. The more important fact in this context is that *condition (2.29) includes the Blum-DeHardt sufficient condition from Theorem 2.1 as a particular case*. This follows readily by noticing that the Blum-DeHardt condition is obtained in (2.29) by requiring that the infimum is attained for $n = 1$.

Our next aim is to show that (2.29) is necessary and sufficient for (2.24) with any of the convergence concepts mentioned above. The first result in this direction may be stated as follows.

Theorem 2.4.

Let $\xi = \{ \xi_j \mid j \geq 1 \}$ *be a stationary ergodic sequence of random variables defined on a probability space* $(\Omega, \mathcal{F}, P)$ *with values in a measurable space* $(S, \mathcal{A})$ *and a common distribution law* π *, let* T *be an arbitrary set, and let* $f : S \times T \to \mathbf{R}$ *be a given map. Let us suppose that* $\| f \|_T$ *belongs to* $L^{<1>}(\pi)$ *and that the map* $s \mapsto f(s,t)$ *is* π*-measurable for every* $t \in T$ *, and let* $M(t) = \int_S f(s,t)\, \pi(ds)$ *be the* π*-mean function of* f *for* $t \in T$ *. If* f *is eventually totally bounded in* π*-mean and* ξ *is* P*-perfect as a map from* Ω *into* $S^{\mathbf{N}}$ *, then we have:*

$$\| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j) - M \|_T \to 0 \quad (a.s.)^* \ \& \ (L^1)^* \tag{2.30}$$

as $n \to \infty$.

Proof. Let $\varepsilon > 0$ be given, then there exists $\gamma_\varepsilon \in \Gamma(T)$ satisfying:

$$\inf_{n \geq 1} \ \frac{1}{n} \int^* \sup_{t',t'' \in A} \mid S_n\big(f(t') - f(t'')\big) \mid \ dP < \varepsilon \tag{2.31}$$

for all $A \in \gamma_\varepsilon$. Since under our hypotheses M belongs to $B(T)$, there is no restriction to assume that for all $A \in \gamma_\varepsilon$ we also have:

$$\sup_{t',t'' \in A} \mid M(t') - M(t'') \mid \ \leq \ \varepsilon \ . \tag{2.32}$$

For every A in γ_ε , choose a point t_A in A . Then by (2.32) we have:

$$\begin{aligned} \| \frac{1}{n} S_n(f) - M \|_T = \ & \max_{A \in \gamma_\varepsilon} \| \frac{1}{n} S_n(f) - M \|_A \leq \max_{A \in \gamma_\varepsilon} \Big\{ \sup_{t',t'' \in A} \mid \frac{1}{n} S_n(f(t')) \\ & - \frac{1}{n} S_n(f(t'')) \mid + \mid \frac{1}{n} S_n(f(t_A)) - M(t_A) \mid \\ & + \sup_{t',t'' \in A} \mid M(t') - M(t'') \mid \Big\} \ \leq \ \max_{A \in \gamma_\varepsilon} \sup_{t',t'' \in A} \mid \frac{1}{n} S_n\big(f(t') - f(t'')\big) \mid \\ & + \max_{A \in \gamma_\varepsilon} \mid \frac{1}{n} S_n(t_A)) - M(t_A) \mid + \ \varepsilon \ . \end{aligned}$$

By the law of large numbers for stationary sequences (2.12) hence we easily get:

$$\begin{aligned} \limsup_{n \to \infty} \ \| \frac{1}{n} S_n(f) - M \|_T^* \leq \ & \limsup_{n \to \infty} \ \max_{A \in \gamma_\varepsilon} \Big(\sup_{t',t'' \in A} \mid \frac{1}{n} S_n\big(f(t') - f(t'')\big) \mid \Big)^* + \ \varepsilon \\ = \ & \max_{A \in \gamma_\varepsilon} \ \limsup_{n \to \infty} \Big(\sup_{t',t'' \in A} \mid \frac{1}{n} S_n\big(f(t') - f(t'')\big) \mid \Big)^* + \ \varepsilon \ . \end{aligned}$$

Therefore by (2.31) we may conclude that in order to establish the $(a.s.)^*$-convergence in (2.30) it is enough to show the following inequality:

$$
\text{(2.33)} \qquad \limsup_{n\to\infty} \Big(\sup_{t',t''\in A} \ | \ \frac{1}{n} S_n\big(f(t')-f(t'')\big) \ | \ \Big)^* \\
\leq \ \inf_{n\geq 1} \int^* \sup_{t',t''\in A} \ | \ \frac{1}{n} S_n\big(f(t')-f(t'')\big) \ | \ dP
$$

for every $A \in \gamma_\varepsilon$. We leave this fact to be established with some additional information in the next proposition. Thus we shall proceed with the $(L^1)^*$-convergence in (2.30).

Let g^* denote the upper π-envelope of the map $s \mapsto \| \ f(s) - M \ \|_T$. By our assumptions we may easily verify that g^* belongs to $L^1(\pi)$. Therefore by (2.10) and (2.13) the sequence of averages $\{ \ \frac{1}{n} \sum_{j=1}^n g^*(\xi_j) \ | \ n \geq 1 \ \}$ is uniformly integrable. Now note that we have:

$$
\| \ \frac{1}{n} S_n(f) - M \ \|_T^* \ \leq \frac{1}{n} \sum_{j=1}^n g^*(\xi_j)
$$

for all $n \geq 1$, and thus the sequence $\big\{ \ \| \ \frac{1}{n} S_n(f) - M \ \|_T^* \ \ | \ \ n \geq 1 \ \big\}$ is uniformly integrable as well. Therefore the $(L^1)^*$-convergence follows straightforwardly by the $(a.s.)^*$-convergence. These facts complete the proof. $\square$

In order to prove the key inequality (2.33) in the preceding proof, and to obtain some additional information, we shall first recall the following fact which is easily verified. Given a stationary sequence $\xi = \{ \ \xi_j \ | \ j \geq 1 \ \}$ on $(\Omega, \mathcal{F}, P)$ with values in $(S, \mathcal{A})$, the unilateral shift θ is P_ξ-perfect in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$. In other words, for any function G from $S^{\mathbf{N}}$ into $\bar{\mathbf{R}}$ we have $(G \circ \theta)^* = G^* \circ \theta$. The next proposition finishes the proof of the preceding theorem.

Proposition 2.5

Under the hypotheses in Theorem 2.4 let us suppose that $\| \ f \ \|_T$ *belongs to* $L^{<1>}(\pi)$ *and that the map* $s \mapsto f(s,t)$ *is* π*-measurable for every* $t \in T$ *. If* ξ *is* P*-perfect as a map from* Ω *into* $S^{\mathbf{N}}$ *, then for any subset* A *of* T *the following three statements are satisfied:*

$$
\text{(2.34)} \qquad \limsup_{n\to\infty} \Big(\sup_{t',t''\in A} \ | \ \frac{1}{n} S_n\big(f(t')-f(t'')\big) \ | \ \Big)^* = \ C \quad P\text{-}a.s.
$$

$$
\text{(2.35)} \qquad C \ \leq \ \inf_{n\geq 1} \int^* \sup_{t',t''\in A} \ | \ \frac{1}{n} S_n\big(f(t')-f(t'')\big) \ | \ dP
$$

$$
\text{(2.36)} \qquad \inf_{n\geq 1} \int^* \sup_{t',t''\in A} \ | \ \frac{1}{n} S_n\big(f(t')-f(t'')\big) \ | \ dP \\
= \ \limsup_{n\to\infty} \int^* \sup_{t',t''\in A} \ | \ \frac{1}{n} S_n\big(f(t')-f(t'')\big) \ | \ dP
$$

where C *is a real constant depending on* A *.*

Proof. There is no restriction to assume that A itself is equal to T , and that the expression $\sup_{t',t''\in A} \ | \ n^{-1} S_n\big(f(t')-f(t'')\big) \ |$ involving the differences $f(t')-f(t'')$ in (2.34)-(2.36) is replaced by the expression $\sup_{t\in A} \mid n^{-1}S_n(f(t)) \mid$ involving the single function $f(t)$. Note that our hypotheses remain valid after this change.

For every $n \geq 1$ define a map $G_n : S^{\mathbf{N}} \to \bar{\mathbf{R}}$ by:

$$G_n(s) \;=\; \sup_{t\in T} \mid \frac{1}{n}\sum_{j=1}^{n} f(s_j,t) \mid \;=\; \| \frac{1}{n}\sum_{j=1}^{n} f(s_j) \|_T$$

for $s = (s_1, s_2, \dots) \in S^{\mathbf{N}}$. For given $d \geq 1$ and $n > d$, put $\sigma_n = [n/d]$ to be the integer part of n/d . Then we have:

$$G_n(s) \;\leq\; \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n}\sum_{j=1}^{\sigma_n} \big(G_d \circ \theta^{(j-1)d}\big)(s) \;+\; \frac{1}{n} \sum_{\sigma_n\cdot d<j\leq n} \| f(s_j) \|_T \tag{2.37}$$

for all $s = (s_1, s_2, \dots) \in S^{\mathbf{N}}$. Indeed, let us note that:

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} f(s_j) \;&=\; \frac{1}{n}\sum_{j=1}^{\sigma_n\cdot d} f(s_j) \;+\; \frac{1}{n}\sum_{\sigma_n\cdot d<j\leq n} f(s_j) \\ &=\; \frac{1}{n}\sum_{j=1}^{\sigma_n}\sum_{i=1}^{d} f(s_{i+(j-1)d}) \;+\; \frac{1}{n}\sum_{\sigma_n\cdot d<j\leq n} f(s_j) \\ &=\; \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n}\sum_{j=1}^{\sigma_n} \Big(\frac{1}{d}\sum_{i=1}^{d} f(s_{i+(j-1)d}) \Big) \;+\; \frac{1}{n}\sum_{\sigma_n\cdot d<j\leq n} f(s_j) \end{aligned}$$

for all $s = (s_1, s_2, \dots) \in S^{\mathbf{N}}$. Therefore (2.37) follows easily by taking the supremum over all $t \in T$. Taking the upper P_ξ-envelopes of G_n and G_d in (2.37) we get:

$$G_n^*(s) \;\leq\; \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n}\sum_{j=1}^{\sigma_n} \big(G_d^* \circ \theta^{(j-1)d}\big)(s) \;+\; \frac{1}{n} \sum_{\sigma_n\cdot d<j\leq n} \| f(s_j) \|_T^*$$

for all $s = (s_1, s_2, \dots) \in S^{\mathbf{N}}$, where $\| f \|_T^*$ denotes the upper π-envelope of $\| f \|_T$ as a function from S into $\bar{\mathbf{R}}$. Hence by P-perfectness of ξ we directly find:

$$\begin{aligned} \| \frac{1}{n}\sum_{j=1}^{n} f(\xi_j) \|_T^* \;=\; G_n^*(\xi) \;\leq\; &\frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n}\sum_{j=1}^{\sigma_n} \big(G_d^* \circ \theta^{(j-1)d}\big)(\xi) \\ &+ \frac{1}{n} \sum_{\sigma_n\cdot d<j\leq n} \| f \|_T^* \circ \xi_j \ . \end{aligned} \tag{2.38}$$

Next note by (2.10) and (2.11) that the sequence $\{ \big(G_d^* \circ \theta^{(j-1)d}\big)(\xi) \mid j \geq 1\}$ is stationary and ergodic. Moreover, since $\| f \|_T$ belongs to $L^{<1>}(\pi)$, then obviously:

$$\int (G_d^* \circ \xi)\; dP \;\leq\; \int^* \| f(s) \|_T \;\; \pi(ds) \;<\; \infty \ .$$

Thus the law of large numbers (2.12) may be applied, and since $(\sigma_n \cdot d)/n \to 1$ as $n \to \infty$, we may conclude:

$$\text{(2.39)} \qquad \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n} \sum_{j=1}^{\sigma_n} \big(G_d^* \circ \theta^{(j-1)d}\big)(\xi) \;\longrightarrow\; \int_\Omega (G_d^* \circ \xi)\; dP \quad P\text{-}a.s.$$

as $n \to \infty$. Since ξ is by our assumption P-perfect, we have:

$$\text{(2.40)} \qquad \int_\Omega (G_d^* \circ \xi)\; dP \;=\; \int_\Omega^* \Big\| \frac{1}{d} \sum_{j=1}^{d} f(\xi_j) \Big\|_T \, dP \; .$$

Similarly, by (2.10) and (2.11) we may conclude that the sequence $\{ \, \| f \|_T^* \circ \xi_j \;\mid\; j \geq 1 \, \}$ is stationary and ergodic. Moreover, since $\| f \|_T$ belongs to $L^{<1>}(\pi)$, then obviously:

$$\int_\Omega^* \| f \|_T^* \circ \xi_1 \; dP \;=\; \int^* \| f(s) \|_T \;\; \pi(ds) \;<\; \infty \; .$$

Thus the law of large numbers (2.12) may be applied, and since $(\sigma_n \cdot d)/n \to 1$ as $n \to \infty$, we may conclude:

$$\text{(2.41)} \qquad \frac{1}{n} \sum_{\sigma_n \cdot d < j \leq n} \| f \|_T^* \circ \xi_j \;=\; \frac{1}{n} \sum_{j=1}^{n} \| f \|_T^* \circ \xi_j \;-\; \frac{1}{n} \sum_{j=1}^{\sigma_n \cdot d} \| f \|_T^* \circ \xi_j$$

$$=\; \frac{1}{n} \sum_{j=1}^{n} \| f \|_T^* \circ \xi_j \;-\; \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n \cdot d} \sum_{j=1}^{\sigma_n \cdot d} \| f \|_T^* \circ \xi_j \;\to\; 0 \quad P\text{-}a.s.$$

as $n \to \infty$. Now by (2.38)-(2.41) we obtain:

$$\text{(2.42)} \qquad \limsup_{n\to\infty} \; \Big\| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j) \Big\|_T^* \;\leq\; \int_\Omega^* \Big\| \frac{1}{d} \sum_{j=1}^{d} f(\xi_j) \Big\|_T \, dP \quad P\text{-}a.s.$$

for all $d \geq 1$. Thus by showing (2.34), the statement (2.35) will also be established.

To prove (2.34), since by our assumption ξ is P-perfect, it is enough to show that we have:

$$\text{(2.43)} \qquad \limsup_{n\to\infty} \; G_n^*(\xi) \;=\; C \quad P\text{-}a.s.$$

In order to establish (2.43) it is enough to show that the map $\limsup_{n\to\infty} \; G_n^*$ is θ-invariant P_ξ-a.s., or in other words (see Paragraph 4 in Section 1.1) that:

$$\text{(2.44)} \qquad \limsup_{n\to\infty} \; G_n^* \circ \theta \;=\; \limsup_{n\to\infty} \; G_n^* \quad P_\xi\text{-}a.s.$$

Since θ is P_ξ-perfect, we have $G_n^* \circ \theta = (G_n \circ \theta)^*$. Moreover, we also have:

$$(G_n \circ \theta) \;=\; \sup_{t \in T} \; \Big| \, \frac{1}{n} \sum_{j=1}^{n} f(s_{j+1}, t) \, \Big| \;=\; \sup_{t \in T} \; \Big| \, \frac{n+1}{n} \cdot \frac{1}{n+1} \sum_{j=1}^{n+1} f(s_j, t) \;-\; \frac{1}{n} \cdot f(s_1, t) \, \Big|$$

for all $s = (s_1, s_2, \ldots\,) \in S^{\mathbf{N}}$. Using these two facts one can easily verify the validity of (2.44). Note since $\| f \|_T$ belongs to $L^{<1>}(\pi)$ that we have $\| f \|_T < \infty$ π-$a.s.$. Thus $\limsup_{n\to\infty} \; G_n^*$ is θ-invariant mod P_ξ , and (2.43) follows from the ergodicity of ξ .

To prove (2.36) take the P-integral on both sides in (2.38). Then by (2.40) we get:

$$\int^* \| \frac{1}{n} \sum_{j=1}^n f(\xi_j) \|_T \, dP \;\leq\; \frac{\sigma_n \cdot d}{n} \int_\Omega^* \| \frac{1}{d} \sum_{j=1}^d f(\xi_j) \|_T \, dP \;+\; \frac{n - \sigma_n \cdot d}{n} \int^* \| f \|_T \; \pi(ds) \; .$$

Since $(\sigma_n \cdot d)/n \leq 1$ and $(n - \sigma_n \cdot d)/n \leq d/n$, and since by our assumption $\| f \|_T$ belongs to $L^{<1>}(\pi)$, we may conclude:

$$\limsup_{n \to \infty} \int^* \| \frac{1}{n} \sum_{j=1}^n f(\xi_j) \|_T \, dP \;\leq\; \int^* \| \frac{1}{d} \sum_{j=1}^d f(\xi_j) \|_T \, dP$$

for all $d \geq 1$. Now (2.36) follows straightforwardly by taking the infimum over all $d \geq 1$. This fact completes the proof. $\square$

Remark 2.6

Under the hypotheses in Theorem 2.4 and Proposition 2.5 it is easily verified that in the case when the map $(s_1, s_2, \dots) \mapsto \sup_{t', t'' \in A} | \frac{1}{n} \sum_{j=1}^n (f(s_j, t') - f(s_j, t'')) |$ is P_ξ-measurable for all $n \geq 1$, the assumption of P-perfectness on ξ is not needed for their conclusions remain valid.

Theorem 2.7

Let $\xi = \{ \xi_j \mid j \geq 1 \}$ be a stationary ergodic sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$ and a common distribution law π , let T be an arbitrary set, and let $f : S \times T \to \mathbf{R}$ be a given map. Let us suppose that $\| f \|_T$ belongs to $L^{<1>}(\pi)$ and that the map $s \mapsto f(s,t)$ is π-measurable for every $t \in T$, and let $M(t) = \int_S f(s,t) \, \pi(ds)$ be the π-mean function of f for $t \in T$. If we have:

$$\| \frac{1}{n} \sum_{j=1}^n f(\xi_j) - M \|_T \to 0 \quad (P^*) \tag{2.45}$$

as $n \to \infty$, then f is eventually totally bounded in π-mean.

Proof. Let us for $n \geq 1$ and $\alpha \in \Gamma(T)$ denote:

$$D_{n,\alpha} = \Big(\max_{A \in \alpha} \sup_{t', t'' \in A} | \frac{1}{n} S_n\big(f(t') - f(t'')\big) | \Big)^* = \max_{A \in \alpha} \Big(\sup_{t', t'' \in A} | \frac{1}{n} S_n\big(f(t') - f(t'')\big) | \Big)^* \; .$$

Then we obviously have:

$$D_{n,\alpha} \;\leq\; 2 \, \| \frac{1}{n} S_n(f) \|_T^* \;\leq\; \frac{2}{n} \sum_{j=1}^n \| f \|_T^* \circ \xi_j$$

for all $n \geq 1$ and all $\alpha \in \Gamma(T)$. Therefore by (2.10) and (2.13) we may conclude that the family of random variables $\{ D_{n,\alpha} \mid n \geq 1 \, , \; \alpha \in \Gamma(T) \}$ is uniformly integrable. Thus for given $\varepsilon > 0$, there exists $0 < \delta < \varepsilon/2$ such that:

$$\int_F D_{n,\alpha} \, dP < \varepsilon/2 \tag{2.46}$$

for all $F \in \mathcal{F}$ satisfying $P(F) < \delta$, whenever $n \geq 1$ and $\alpha \in \Gamma(T)$. Since under our assumptions M belongs to $B(T)$, then for given $\delta > 0$, there exists $\alpha_\delta \in \Gamma(T)$ such that:

$$\sup_{t',t'' \in A} \mid M(t') - M(t'') \mid \le \delta/2 \tag{2.47}$$

for all $A \in \gamma_\delta$. Thus by (2.45) and (2.47) we easily get:

$$\begin{aligned} P\{ D_{n,\gamma_\delta} > \delta \} \ &\le \ P^*\{ 2 \parallel \frac{1}{n} S_n(f) - M \parallel_T \\ &\quad + \max_{A \in \gamma_\delta} \sup_{t',t'' \in A} \mid M(t') - M(t'') \mid > \delta \} \\ &\le \ P^*\{ \parallel \frac{1}{n} S_n(f) - M \parallel_T > \delta/4 \} < \delta \end{aligned} \tag{2.48}$$

for all $n \ge n_\delta$ with some $n_\delta \ge 1$. Now by (2.46) and (2.48) we may conclude:

$$\int D_{n,\gamma_\delta} \, dP \ = \ \int_{\{D_{n,\gamma_\delta} \le \delta\}} D_{n,\gamma_\delta} \, dP \ + \ \int_{\{D_{n,\gamma_\delta} > \delta\}} D_{n,\gamma_\delta} \, dP \ \le \ \delta + \varepsilon/2 \ < \varepsilon$$

for all $n \ge n_\delta$. Hence we get:

$$\inf_{n \ge 1} \int^* \sup_{t',t'' \in A} \mid \frac{1}{n} S_n\big(f(t') - f(t'')\big) \mid \ dP \ \le \ \limsup_{n \to \infty} \int D_{n,\gamma_\delta} \, dP < \varepsilon$$

for all $A \in \gamma_\delta$. Thus f is eventually totally bounded in π-mean, and the proof is complete. □

Corollary 2.8

Under the hypotheses in Theorem 2.7 let us suppose that $\parallel f \parallel_T$ belongs to $L^{<1>}(\pi)$ and that the map $s \mapsto f(s,t)$ is π-measurable for every $t \in T$, and let $M(t) = \int_S f(s,t) \, \pi(ds)$ be the π-mean function of f for $t \in T$. If ξ is P-perfect as a map from Ω into $S^{\mathbf{N}}$, then the following four statements are equivalent:

$$\Big\| \frac{1}{n} S_n(f) - M \Big\|_T \to 0 \ \ (a.s.)^* \tag{2.49}$$

$$\Big\| \frac{1}{n} S_n(f) - M \Big\|_T \to 0 \ \ (L^1)^* \tag{2.50}$$

$$\Big\| \frac{1}{n} S_n(f) - M \Big\|_T \to 0 \ \ (P^*) \tag{2.51}$$

(2.52) *The map f is eventually totally bounded in π-mean.*

Proof. Straightforward from Theorem 2.4 and Theorem 2.7 by using (2.22). □

The next proposition shows that the $(L^1)^*$-convergence and the (P^*)-convergence in the preceding corollary may be seemingly relaxed.

Proposition 2.9

Under the hypotheses in Theorem 2.7 let us suppose that $\parallel f \parallel_T$ belongs to $L^{<1>}(\pi)$ and that the map $s \mapsto f(s,t)$ is π-measurable for every $t \in T$, and let $M(t) = \int_S f(s,t) \, \pi(ds)$ be *the π-mean function of f for $t \in T$. If ξ is P-perfect as a map from Ω into $S^{\mathbf{N}}$, then we have $\| \frac{1}{n} S_n(f) - M \|_T \to 0 \ (P^*)$, if and only if we have:*

$$\inf_{n \geq 1} P^* \{ \| \frac{1}{n} S_n(f) - M \|_T > \varepsilon \} < \varepsilon \tag{2.53}$$

for every $\varepsilon > 0$. Similarly, if ξ is P-perfect as a map from Ω into $S^{\mathbf{N}}$, then we have $\| \frac{1}{n} S_n(f) - M \|_T \to 0 \ (L^1)^*$, if and only if we have:

$$\inf_{n \geq 1} \int^* \| \frac{1}{n} S_n(f) - M \|_T \, dP \ = \ 0 \ . \tag{2.54}$$

Proof. First we show that (2.53) implies (2.54). So, suppose that (2.53) holds, and denote:

$$D_n \ = \ \| \frac{1}{n} S_n(f) - M \|_T^*$$

for all $n \geq 1$. Then we have:

$$D_n \ \leq \ \frac{1}{n} \sum_{j=1}^{n} \| f \|_T^* \circ \xi_j + \ \| M \|_T$$

for all $n \geq 1$. Since by our hypotheses $\| f \|_T$ belongs to $L^{<1>}(\pi)$ and M belongs to $B(T)$, then by (2.10) and (2.13) the sequence of random variables $\{ D_n \mid n \geq 1 \}$ is uniformly integrable. Thus for given $\varepsilon > 0$, there exists $0 < \delta < \varepsilon/2$ such that:

$$\int_F D_n \, dP < \varepsilon/2 \tag{2.55}$$

whenever $F \in \mathcal{F}$ with $P(F) < \delta$, for all $n \geq 1$. Given this $\delta > 0$ we know by (2.53) that there is $n_\delta \geq 1$ such that:

$$P\{ D_{n_\delta} > \delta \} < \delta \ . \tag{2.56}$$

Now by (2.55) and (2.56) we may conclude:

$$\int^* \| \frac{1}{n_\delta} S_{n_\delta}(f) - M \|_T \ dP \ = \ \int D_{n_\delta} \, dP \ = \ \int_{\{D_{n_\delta} \leq \delta\}} D_{n_\delta} \, dP$$

$$+ \int_{\{D_{n_\delta} > \delta\}} D_{n_\delta} \, dP \ \leq \ \delta \ + \ \varepsilon/2 \ < \ \varepsilon \ .$$

This fact establishes (2.54). To show that (2.54) implies the $(L^1)^*$-convergence one can easily verify that the same arguments as the ones given in the proof of Proposition 2.5 apply here as well, and in this way one obtains:

$$\inf_{n \geq 1} \int^* \| \frac{1}{n} S_n(f) - M \|_T \ dP \ = \ \limsup_{n \to \infty} \int^* \| \frac{1}{n} S_n(f) - M \|_T \ dP \ .$$

Therefore (2.54) implies $\| \frac{1}{n} S_n(f) - M \|_T^* \to 0 \ \ (L^1)^*$ as $n \to \infty$. Since by (2.22) the $(L^1)^*$-convergence implies (P^*)-convergence in general, these facts complete the proof. $\square$

We pass to the connection with a weak convergence concept. Let us consider a Banach space B and a sequence of arbitrary functions $\{ Z_j \mid j \geq 1 \}$ from a probability space $(\Omega, \mathcal{F}, P)$ into B . Let $C(B)$ denote the set of all bounded continuous functions from B into $\mathbf{R}$, and let $\mathcal{K}(B)$ denote the family of all compact subsets of B . Let μ be a probability measure defined on the Borel σ-algebra in B , then the sequence $\{ Z_j \mid j \geq 1 \}$ is said to be:

(2.57) *weakly convergent* to μ , if we have:

$$\lim_{n\to\infty} \int^* F(Z_n)\, dP \;=\; \lim_{n\to\infty} \int_* F(Z_n)\, dP \;=\; \int_B F\, d\mu$$

for all $F \in C(B)$, and in this case we shall write $Z_n \to \mu$ *weakly in* B

(2.58) *uniformly tight*, if $\forall \varepsilon > 0$, $\exists K_\varepsilon \in \mathcal{K}(B)$ such that:

$$\limsup_{n\to\infty} P^*\{ Z_n \notin K_\varepsilon \} \leq \varepsilon$$

(2.59) *eventually tight*, if $\forall \varepsilon > 0$, $\exists K_\varepsilon \in \mathcal{K}(B)$ such that:

$$\limsup_{n\to\infty} \int^* F(Z_n)\, dP \leq \varepsilon$$

for all $F \in C(B)$ satisfying $0 \leq F \leq 1_{B\setminus K_\varepsilon}$.

It is easily verified that we have (see [42]):

(2.60) If $\{ Z_j \mid j \geq 1 \}$ is uniformly tight, it also is eventually tight.

(2.61) If $Z_n \to \mu$ weakly in B and μ is Radon, then $\{ Z_j \mid j \geq 1 \}$ is eventually tight.

(2.62) If $Z_n \to c$ (P^*) for some $c \in B$, then $Z_n \to c$ weakly in B .

Clarify that $Z_n \to c$ (P^*) means $\| Z_n - c \| \to 0$ (P^*) where $\| \cdot \|$ denotes the norm in B , and that $Z_n \to c$ weakly in B means $Z_n \to \delta_c$ where δ_c denotes the Dirac measure at c .

Our next aim is to show that in the stationary ergodic case the eventually tightness of the sequence of averages $\{ \frac{1}{n} S_n(f) \mid n \geq 1 \}$ in the Banach space $(B(T), \| \cdot \|_T)$ is equivalent to the uniform law of large numbers. One implication in this direction is obvious. Namely, if $\| \frac{1}{n} S_n(f) - M \|_T \to 0$ (P^*) as $n \to \infty$, then by (2.61) and (2.62) we see that the sequence $\{ \frac{1}{n} S_n(f) \mid n \geq 1 \}$ is eventually tight. The remaining part is contained in the following theorem.

Theorem 2.10

Let $\xi = \{ \xi_j \mid j \geq 1 \}$ be a stationary ergodic sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$ and a common distribution law π , let T be an arbitrary set, and let $f : S \times T \to \mathbf{R}$ be a given map. Let us suppose that $\| f \|_T$ belongs to $L^{<1>}(\pi)$ and that the map $s \mapsto f(s,t)$ is π-measurable for every $t \in T$. If ξ is P-perfect as a map from Ω into $S^{\mathbf{N}}$, then the following two statements are equivalent:

(2.63) *The map f is eventually totally bounded in π-mean*

(2.64) *The sequence of averages $\{ n^{-1} S_n(f) \mid n \geq 1 \}$ is eventually tight in $B(T)$.*

Proof. The implication (2.63) ⇒(2.64) follows by Theorem 2.4, (2.22) and (2.61)+(2.62) as already mentioned. For the converse, suppose that (2.64) holds. Then we claim that for given $\delta > 0$, there exists $\gamma_\delta \in \Gamma(T)$ such that:

$$P^*\{ \max_{A \in \gamma_\delta} \sup_{t',t'' \in A} \big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \big| > \delta \} < \delta \tag{2.65}$$

for all $n \geq n_\delta$ with some $n_\delta \geq 1$. To prove (2.65) we may proceed as follows. Since the sequence of averages $\{ \frac{1}{n} S_n(f) \mid n \geq 1 \}$ is eventually tight in $B(T)$, then for given $\delta > 0$, there exists $K_\delta \in \mathcal{K}(B(T))$ such that:

$$\limsup_{n \to \infty} \int^* F\big(n^{-1} S_n(f)\big) \, dP \ \leq \ \delta \tag{2.66}$$

for all $F \in C(B(T))$ satisfying $0 \leq F \leq 1_{B(T) \setminus K_\delta}$. The compactness of K_δ yields the existence of $\gamma_\delta \in \Gamma(T)$ satisfying:

$$\sup_{t',t'' \in A} | \varphi(t') - \varphi(t'') | < \delta/3 \tag{2.67}$$

for all $\varphi \in K_\delta$ and all $A \in \gamma_\delta$ (see [17] p.260). By using (2.67) one can easily verify that for any $\psi \in b(K_\delta, \delta/3) = \bigcup_{\varphi \in K_\delta} b(\varphi, \delta/3)$ we have:

$$\sup_{t',t'' \in A} | \psi(t') - \psi(t'') | < \delta \tag{2.68}$$

for all $A \in \gamma_\delta$. Now by (2.66) and (2.68) we may conclude:

$$\begin{aligned} &P^*\{ \max_{A \in \gamma_\delta} \sup_{t',t'' \in A} \big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \big| > \delta \} \\ &\quad = P^*\{ \max_{A \in \gamma_\delta} \sup_{t',t'' \in A} \big| \frac{1}{n} S_n\big(f(t')\big) - \frac{1}{n} S_n\big(f(t'')\big) \big| > \delta \} \\ &\quad \leq \int^* 1_{\{n^{-1} S_n(f) \notin b(K_\delta, \delta/3)\}} \, dP \ \leq \ \int^* F_\delta\big(n^{-1} S_n(f)\big) \, dP \ \leq \ \delta \end{aligned}$$

for all $n \geq n_\delta$ with some $n_\delta \geq 1$, where $F_\delta \in C(B(T))$ is chosen to satisfy $F_\delta(\varphi) = 1$ for $\varphi \in B(T) \setminus b(K_\delta, \delta/3)$, $F_\delta(\varphi) = 0$ for $\varphi \in K_\delta$, and $0 \leq F_\delta(\varphi) \leq 1_{B(T) \setminus K_\delta}(\varphi)$ for all $\varphi \in B(T)$. These facts complete the proof of (2.65). Now, denote:

$$D_{n,\alpha} \ = \ \Big(\max_{A \in \alpha} \sup_{t',t'' \in A} \big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \big| \Big)^* = \ \max_{A \in \alpha} \Big(\sup_{t',t'' \in A} \big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \big| \Big)^* .$$

for $n \geq 1$ and $\alpha \in \Gamma(T)$. Then by (2.10) and (2.13) we may conclude as in the proof of Theorem 2.7 that the family $\{ D_{n,\alpha} \mid n \geq 1 \, , \, \alpha \in \Gamma(T) \}$ is uniformly integrable. Therefore for given $\varepsilon > 0$, there exists $0 < \delta < \varepsilon/2$ such that:

$$\int_F D_{n,\alpha} \, dP \ < \ \varepsilon/2 \tag{2.69}$$

whenever $F \in \mathcal{F}$ with $P(F) < \delta$, for all $n \geq 1$ and all $\alpha \in \Gamma(T)$. If we now apply (2.65) with this $\delta > 0$, we can find $\gamma_\delta \in \Gamma(T)$ satisfying:

(2.70) $$P\{\ D_{n,\gamma_\delta} > \delta\ \} < \delta\ .$$

By (2.69) and (2.70) we get:

$$\int D_{n,\gamma_\delta}\ dP\ =\ \int_{\{D_{n,\gamma_\delta}\le\delta\}} D_{n,\gamma_\delta}\ dP\ +\ \int_{\{D_{n,\gamma_\delta}>\delta\}} D_{n,\gamma_\delta}\ dP\ \le\ \delta+\varepsilon/2\ <\varepsilon$$

for all $n \ge 1$. Therefore we may conclude:

(2.71) $$\inf_{n\ge 1} \int^* \sup_{t',t''\in A} \Big|\ \frac{1}{n} S_n\big(f(t') - f(t'')\big)\ \Big|\ dP\ \le\ \limsup_{n\to\infty} \int D_{n,\gamma_\delta}\ dP < \varepsilon$$

for all $A \in \gamma_\delta$, and the proof is complete. □

Corollary 2.11

Under the hypotheses in Corollary 2.8 let us suppose that $\|\ f\ \|_T$ *belongs to* $L^{<1>}(\pi)$ *and that the map* $s \mapsto f(s,t)$ *is* π*-measurable for every* $t \in T$ *, and let* $M(t) = \int_S f(s,t)\ \pi(ds)$ *be the* π*-mean function of* f *for* $t \in T$ *. If* ξ *is* P*-perfect as a map from* Ω *into* $S^{\mathbf{N}}$ *, then the statements (2.49)-(2.52) are also equivalent to the following statement:*

(2.72) $$\frac{1}{n} S_n(f) \to M \quad \text{weakly in} \quad B(T)$$

as $n \to \infty$.

Proof. Straightforward by Corollary 2.8, Theorem 2.10 and (2.61)+(2.62). □

We proceed by considering the property of being eventually totally bounded in the mean in some more technical details.

Theorem 2.12

Under the hypotheses in Theorem 2.10 let us suppose that $\|\ f\ \|_T$ *belongs to* $L^{<1>}(\pi)$ *and that the map* $s \mapsto f(s,t)$ *is* π*-measurable for every* $t \in T$ *. If* ξ *is* P*-perfect as a map from* Ω *into* $S^{\mathbf{N}}$ *, then the following two statements are equivalent:*

(2.73) *The map* f *is eventually totally bounded in* π*-mean*

(2.74) *For every* $\varepsilon > 0$ *, there exists* $\gamma_\varepsilon \in \Gamma(T)$ *such that any of the following seven conditions is satisfied:*

(2.74.1) $$\inf_{n\ge 1} \int^* \max_{A\in\gamma_\varepsilon} \sup_{t',t''\in A} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big)\Big|\ dP\ < \varepsilon$$

(2.74.2) $$\limsup_{n\to\infty} \int^* \sup_{t',t''\in A} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big)\Big|\ dP\ < \varepsilon\ \ ,\ \ \forall A \in \gamma_\varepsilon$$

(2.74.3) $$\limsup_{n\to\infty} \int^* \max_{A\in\gamma_\varepsilon} \sup_{t',t''\in A} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big)\Big|\ dP\ < \varepsilon$$

(2.74.4) $$\inf_{n\ge 1}\ P^*\Big\{ \sup_{t',t''\in A} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big)\Big| > \varepsilon\ \Big\} <\ \varepsilon\ \ ,\ \ \forall A \in \gamma_\varepsilon$$

(2.74.5) $$\inf_{n\ge 1}\ P^*\Big\{ \max_{A\in\gamma_\varepsilon} \sup_{t',t''\in A} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big)\Big| > \varepsilon\ \Big\} <\ \varepsilon$$

$$(2.74.6) \qquad \limsup_{n\to\infty} \; P^*\Big\{ \sup_{t',t''\in A} \Big| \frac{1}{n} S_n\big(f(t')-f(t'')\big)\Big| > \varepsilon \Big\} < \varepsilon \;\; , \;\; \forall A \in \gamma_\varepsilon$$

$$(2.74.7) \qquad \limsup_{n\to\infty} \; P^*\Big\{ \max_{A\in\gamma_\varepsilon} \sup_{t',t''\in A} \Big| \frac{1}{n} S_n\big(f(t')-f(t'')\big)\Big| > \varepsilon \Big\} < \varepsilon \; .$$

Proof. The implication (2.73) $\Rightarrow$ (2.74.7) is established in the proof of Theorem 2.10, see (2.65), as well as the implication (2.73)$\Rightarrow$(2.74.3), see (2.71). To complete the proof one can easily verify by combining deduced and obvious implications that it is enough to show (2.74.4)$\Rightarrow$(2.73). (Note that Markov's inequality for upper integrals (see [62]) may be useful for this verification.) So, suppose that (2.74.4) holds. Let us for $n \geq 1$, $\alpha \in \Gamma(T)$ and $A \in \Gamma(T)$ denote:

$$D_{n,A} \; = \; \Big(\sup_{t',t''\in A} \Big| \frac{1}{n} S_n\big(f(t')-f(t'')\big)\Big| \Big)^* \; .$$

Since we have:

$$D_{n,A} \; \leq \; 2 \Big\| \frac{1}{n} S_n(f) \Big\|_T^* \; \leq \; \frac{2}{n} \sum_{j=1}^n \| \, f \, \|_T^* \circ \xi_j$$

for all $n \geq 1$ and all $A \in \alpha$ with $\alpha \in \Gamma(T)$, then by (2.10) and (2.13) we may conclude that the family of random variables $\{ \, D_{n,A} \mid n \geq 1 \; , \; A \in \alpha \; , \; \alpha \in \Gamma(T) \, \}$ is uniformly integrable. Therefore for given $\varepsilon > 0$, there exists $0 < \delta < \varepsilon/2$ such that:

$$(2.75) \qquad \int_F D_{n,A} \; dP \; < \varepsilon/2$$

whenever $F \in \mathcal{F}$ with $P(F) < \delta$, for all $n \geq 1$ and all $A \in \alpha$ with $\alpha \in \Gamma(T)$. Applying (2.74.4) with this $\delta > 0$, we see that there are $\gamma_\delta \in \Gamma(T)$ and $n_\delta \geq 1$ such that:

$$(2.76) \qquad P\{ \, D_{n_\delta,A} > \delta \, \} < \delta$$

for all $A \in \gamma_\delta$. Now by (2.75) and (2.76) we may conclude:

$$\int^* \sup_{t',t''\in A} \Big| \frac{1}{n_\delta} S_{n_\delta}\big(f(t')-f(t'')\big)\Big| \; dP \; = \; \int D_{n_\delta,A} \; dP$$
$$= \int_{\{D_{n_\delta,A}\leq\delta\}} D_{n_\delta,A} \; dP \; + \; \int_{\{D_{n_\delta,A}>\delta\}} D_{n_\delta,A} \; dP \; \leq \; \delta + \varepsilon/2 \; < \varepsilon$$

for all $A \in \gamma_\delta$. This fact establishes (2.73), and the proof is complete. $\square$

We conclude this section with a characterization of the eventually totally bounded in the mean property in terms of the existence of some totally bounded pseudo-metrics. Recall that a pseudo-metric ρ on a set T is said to be *totally bounded*, if T may be covered by finitely many ρ-balls of any given radius $r > 0$. A pseudo-metric ρ on a set T is called an *ultra-pseudo-metric*, if $\rho(s,t) \leq \rho(s,u) \vee \rho(u,t)$ whenever $s, t, u \in T$.

Theorem 2.13

Let $\xi = \{ \, \xi_j \mid j \geq 1 \, \}$ be a stationary ergodic sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$ and a common distribution law π , let T be an arbitrary set, and let $f : S \times T \to \mathbf{R}$ be a given map. Let us suppose that

$\| f \|_T$ belongs to $L^{<1>}(\pi)$ and that the map $s \mapsto f(s,t)$ is π-measurable for every $t \in T$, and let $M(t) = \int_S f(s,t)\, \pi(ds)$ be the π-mean function of f for $t \in T$. If ξ is P-perfect as a map from Ω into $S^{\mathbf{N}}$, then the following three statements are equivalent:

(2.77) *The map f is eventually totally bounded in π-mean*

(2.78) *There exists a totally bounded ultra-pseudo-metric ρ on T such that either of the following two equivalent conditions is satisfied:*

(2.78.1) $$\lim_{r \downarrow 0} \limsup_{n \to \infty} \int^* \sup_{\rho(t',t'')<r} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \Big| \; dP \; = \; 0$$

(2.78.2) $$\lim_{r \downarrow 0} \inf_{n \geq 1} \int^* \sup_{\rho(t',t'')<r} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \Big| \; dP \; = \; 0$$

(2.79) *For every $\varepsilon > 0$ there exist a totally bounded pseudo-metric ρ_ε on T and $r_\varepsilon > 0$ such that either of the following two equivalent conditions is satisfied:*

(2.79.1) $$\limsup_{n \to \infty} P^*\Big\{ \sup_{\rho_\varepsilon(s,t)<r_\varepsilon} \Big| \frac{1}{n} S_n\big(f(s) - f(t)\big) \Big| > \varepsilon \Big\} < \varepsilon \;\; , \quad \forall t \in T$$

(2.79.2) $$\inf_{n \geq 1} P^*\Big\{ \sup_{\rho_\varepsilon(s,t)<r_\varepsilon} \Big| \frac{1}{n} S_n\big(f(s) - f(t)\big) \Big| > \varepsilon \Big\} < \varepsilon \;\; , \quad \forall t \in T \; .$$

Proof. We first prove that (2.77) implies (2.78.1). So, suppose that (2.77) holds. Let $\{ r_n \mid n \geq 0 \}$ be a sequence of real numbers satisfying $1 = r_0 > r_1 > r_2 > \ \ldots \ > 0$ with $\lim_{n \to \infty} r_n = 0$. Then for every $n \geq 1$, there exists $\gamma_n \in \Gamma(T)$ such that $| M(t') - M(t'') | < (r_{n-1})^2$ for all $t', t'' \in A$ and all $A \in \gamma_n$. There is no restriction to assume that $\gamma_1 \subset \gamma_2 \subset \ \ldots$. Let us define:

$$\rho(t',t'') \; = \; \sup_{n \geq 1} \Big(\; r_{n-1} \cdot \max_{A \in \gamma_n} \; | 1_A(t') - 1_A(t'') | \; \Big)$$

for all $t', t'' \in T$. Then ρ is evidently a totally bounded ultra-pseudo-metric on T , and it is easy to verify that $\rho(t',t'') < \varepsilon$ implies $| M(t') - M(t'') | < \varepsilon^2$ for $0 < \varepsilon < 1$ and $t', t'' \in T$. Therefore by Markov's inequality for upper integrals (see [62]) we may conclude:

$$P^*\Big\{ \sup_{\rho(t',t'')<\varepsilon} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \Big| > \varepsilon \Big\} \; \leq \; \frac{1}{\varepsilon} \int^* \sup_{\rho(t',t'')<\varepsilon} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \Big| \; dP$$
$$\leq \; \frac{1}{\varepsilon} \int^* \sup_{\rho(t',t'')<\varepsilon} \Big(\; \Big| \frac{1}{n} S_n\big(f(t')\big) - M(t') \Big| \; + \; | M(t') - M(t'') |$$
$$+ \; \Big| M(t'') - \frac{1}{n} S_n\big(f(t'')\big) \Big| \; \Big) \, dP \; \leq \; \frac{1}{\varepsilon} \Big(\; 2 \int^* \Big\| \frac{1}{n} S_n(f) - M \Big\|_T \; dP \; + \; \varepsilon^2 \; \Big)$$

for all $0 < \varepsilon < 1$ and all $n \geq 1$. Hence by (2.77) and Theorem 2.4 we obtain:

(2.80) $$\limsup_{n \to \infty} P^*\Big\{ \sup_{\rho(t',t'')<\varepsilon} \Big| \frac{1}{n} S_n\big(f(t') - f(t'')\big) \Big| > \varepsilon \Big\} \leq \varepsilon$$

for all $0 < \varepsilon < 1$. Let us denote:

$$D_{n,\varepsilon} \;=\; \Big(\sup_{\rho(t',t'')<\varepsilon} \mid \frac{1}{n} S_n\big(f(t') - f(t'')\big) \mid \Big)^*$$

for all $n \geq 1$ and all $\varepsilon > 0$. Then we have:

$$D_{n,\varepsilon} \;\leq\; \frac{2}{n} \sum_{j=1}^{n} \parallel f \parallel_T^* \circ \xi_j$$

for all $n \geq 1$ and all $\varepsilon > 0$. Therefore by (2.10) and (2.13) the family of random variables $\{ D_{n,\varepsilon} \mid n \geq 1 \ , \ \varepsilon > 0 \}$ is uniformly integrable. Thus for given $\varepsilon > 0$, there exists $0 < \delta < (\varepsilon/2) \wedge 1$ such that:

$$\int_F D_{n,\delta} \; dP \;\leq\; \varepsilon/2 \tag{2.81}$$

whenever $F \in \mathcal{F}$ with $P(F) < \delta$. Now by (2.80) and (2.81) we get:

$$\int D_{n,\delta} \; dP \;=\; \int_{\{D_{n,\delta} \leq \delta\}} D_{n,\delta} \; dP \;+\; \int_{\{D_{n,\delta} > \delta\}} D_{n,\delta} \; dP \;\leq\; \delta + \varepsilon/2 \;<\varepsilon$$

for all $n \geq n_\delta$ with some $n_\delta \geq 1$. Hence we obtain:

$$\limsup_{n\to\infty} \int^* \sup_{\rho(t',t'')<\delta} \mid \frac{1}{n} S_n\big(f(t') - f(t'')\big) \mid \; dP \;=\; \limsup_{n\to\infty} \int D_{n,\delta} \; dP \;<\; \varepsilon$$

and (2.78.1) is proved. Using the same arguments as in the proof of Proposition 2.5 we easily get:

$$\limsup_{n\to\infty} \int^* \sup_{\rho(t',t'')<r} \mid \frac{1}{n} S_n\big(f(t') - f(t'')\big) \mid \; dP$$
$$=\; \inf_{n\geq 1} \int^* \sup_{\rho(t',t'')<r} \mid \frac{1}{n} S_n\big(f(t') - f(t'')\big) \mid \; dP$$

for all $r > 0$. Therefore the equivalence between (2.78.1) and (2.78.2) is obvious. The implications (2.78.1) $\Rightarrow$ (2.79.1) and (2.78.2) $\Rightarrow$ (2.79.2) follow by applying Markov's inequality for upper integrals (see [62]). Since (2.79.1) implies (2.79.2) obviously, it is enough to show that (2.79.2) implies (2.77). So, suppose that (2.79.2) holds. Then for given $\varepsilon > 0$ there exists a totally bounded pseudo-metric ρ_ε on T and $r_\varepsilon > 0$ satisfying (2.79.2). Since ρ_ε is totally bounded we can cover the whole set T by finitely many ρ-balls $B_1, B_2, \ldots, B_n$ of given radius r_ε . Therefore putting $\gamma_\varepsilon = \{ B_1, B_2, \ldots, B_n \}$ we see that (2.74.4) is satisfied (with 2ε instead of ε) and in this way we obtain (2.77). These facts complete the proof. $\square$

2.3 The uniform ergodic lemma

1. The proof of the uniform ergodic theorem in the next section relies upon a fact of independent interest that is presented in Theorem 2.14 below. It is instructive to compare this result with the Maximal Ergodic Lemma 1.4, and their roles in the proofs of the ergodic theorems. In Corollary 2.15 below we present its consequence which should be compared with the Maximal Ergodic Inequality

1.5. We begin by introducing the notation and recalling some facts from the non-measurable calculus needed in the sequel. For more details in this direction we refer to [1] and [62].

2. Given a linear operator T in $L^1(\mu)$ and $f \in L^1(\mu)$, we will denote:

(2.82) $$S_n(f) = \sum_{j=0}^{n-1} T^j f$$

(2.83) $$M_n(f) = \max_{1 \leq j \leq n} S_j(f)$$

(2.84) $$R_n(f) = \max_{1 \leq j \leq n} S_j(f)/j$$

for $n \geq 1$. The operator T is said to be *positive*, if $Tf \geq 0$ whenever $f \geq 0$. The operator T is said to be a *contraction* in $L^1(\mu)$, if $\int |Tf| \, d\mu \leq \int |f| \, d\mu$ for all $f \in L^1(\mu)$. We shall restrict our attention to the case where the underlying measure space $(X, \mathcal{A}, \mu)$ is σ-finite, but we remark that further extensions are possible. The symbols μ^* and μ_* denote the *outer μ-measure* and the *inner μ-measure* respectively. The *upper μ-integral* of an arbitrary function f from X into $\bar{\mathbf{R}}$ is defined as follows $\int^* f \, d\mu = \inf \{ \int g \, d\mu \mid g \in L^1(\mu) \, , f \leq g \}$, with the convention $\inf \emptyset = +\infty$. The *lower μ-integral* of an arbitrary function f from X into $\bar{\mathbf{R}}$ is defined as follows $\int_* f \, d\mu = \sup \{ \int g \, d\mu \mid g \in L^1(\mu) \, , \, g \leq f \}$, with the convention $\sup \emptyset = -\infty$. We denote by f^* the *upper μ-envelope* of f . This means that f^* is an $\mathcal{A}$-measurable function from X into $\bar{\mathbf{R}}$ satisfying $f \leq f^*$, and if g is another $\mathcal{A}$-measurable function from X into $\bar{\mathbf{R}}$ satisfying $f \leq g$ μ-a.a., then $f^* \leq g$ μ-a.a. We denote by f_* the *lower μ-envelope* of f . This means that f_* is an $\mathcal{A}$-measurable function from X into $\bar{\mathbf{R}}$ satisfying $f_* \leq f$, and if g is another $\mathcal{A}$-measurable function from X into $\bar{\mathbf{R}}$ satisfying $g \leq f$ μ-a.a., then $g \leq f_*$ μ-a.a. It should be noted that such envelopes exist under the assumption that μ is σ-finite. Recall that we have $\int^* f \, d\mu = \int f^* \, d\mu$, whenever the integral on the right-hand side exists in $\bar{\mathbf{R}}$, and $\int^* f \, d\mu = +\infty$ otherwise. Similarly we have $\int_* f \, d\mu = \int f_* \, d\mu$, whenever the integral on the right-hand side exists in $\bar{\mathbf{R}}$, and $\int_* f \, d\mu = -\infty$ otherwise. To conclude the preliminary part of the section, we clarify that $\int_A^* f \, d\mu$ stands for $\int^* f \cdot 1_A \, d\mu$.

Theorem 2.14 (The uniform ergodic lemma 1993)

Let T be a positive contraction in $L^1(\mu)$, and let $\{ f_\lambda \mid \lambda \in \Lambda \}$ be a family of functions from $L^1(\mu)$. Let us denote $A_n = \{ \sup_{\lambda \in \Lambda} M_n(f_\lambda) > 0 \}$ and $B_n = \{ (\sup_{\lambda \in \Lambda} M_n(f_\lambda))^* > 0 \}$ for all $n \geq 1$. Then we have:

(2.85) $$\int_{A_n}^* \sup_{\lambda \in \Lambda} f_\lambda \, d\mu \geq 0$$

(2.86) $$\int_{B_n}^* \sup_{\lambda \in \Lambda} f_\lambda \, d\mu \geq 0$$

for all $n \geq 1$.

Proof. We shall first assume that $\sup_{\lambda \in \Lambda} f_\lambda \leq g$ for some $g \in L^1(\mu)$. In this case we have $(\sup_{\lambda \in \Lambda} f_\lambda)^* \in L^1(\mu)$. Let $n \geq 1$ be given and fixed. By the monotonicity of T we get:

(2.87) $$S_j(f_\lambda) = f_\lambda + T S_{j-1}(f_\lambda) \leq f_\lambda + T((M_n(f_\lambda))^+)$$

for all $j = 2, \ldots, n+1$, and all $\lambda \in \Lambda$. Moreover, since $T((M_n(f_\lambda))^+) \geq 0$ for all $\lambda \in \Lambda$, we see that (2.87) is valid for $j = 1$ as well. Hence we find:

$$M_n(f_\lambda) \leq f_\lambda + T((M_n(f_\lambda))^+)$$

for all $\lambda \in \Lambda$. Taking the supremum over all $\lambda \in \Lambda$ we obtain:

$$\sup_{\lambda \in \Lambda} M_n(f_\lambda) \leq \sup_{\lambda \in \Lambda} f_\lambda + \sup_{\lambda \in \Lambda} T((M_n(f_\lambda))^+) \ . \tag{2.88}$$

Since $f_\lambda \leq g$ for all $\lambda \in \Lambda$, we have $\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+ \leq (M_n(g))^+$. Hence we see that $(\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+)^* \in L^1(\mu)$. Therefore by (2.88) and the monotonicity of T we get:

$$\sup_{\lambda \in \Lambda} M_n(f_\lambda) \leq \sup_{\lambda \in \Lambda} f_\lambda + T\Big(\big(\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+\big)^*\Big) \ . \tag{2.89}$$

Multiplying both sides by 1_{A_n} we obtain:

$$\begin{aligned} &\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+ = \Big(\sup_{\lambda \in \Lambda} M_n(f_\lambda))\Big)^+ = \sup_{\lambda \in \Lambda} M_n(f_\lambda) \cdot 1_{A_n} \\ &\leq \sup_{\lambda \in \Lambda} f_\lambda \cdot 1_{A_n} + T\Big(\big(\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+\big)^*\Big) \cdot 1_{A_n} \\ &\leq \sup_{\lambda \in \Lambda} f_\lambda \cdot 1_{A_n} + T\Big(\big(\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+\big)^*\Big) \ . \end{aligned}$$

Integrating both sides we get:

$$\begin{aligned} &\int^* \sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+ \, d\mu \\ &\leq \int_{A_n}^* \sup_{\lambda \in \Lambda} f_\lambda \, d\mu + \int T((\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+)^*) \, d\mu \ . \end{aligned}$$

This can be rewritten as follows:

$$\int^* \sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+ \, d\mu - \int T((\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+)^*) \, d\mu \leq \int_{A_n}^* \sup_{\lambda \in \Lambda} f_\lambda \, d\mu$$

and the proof of (2.85) follows from the contractibility of T .

Similarly, from (2.89) we get:

$$\Big(\sup_{\lambda \in \Lambda} M_n(f_\lambda)\Big)^* \leq \Big(\sup_{\lambda \in \Lambda} f_\lambda\Big)^* + T\Big(\big(\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+\big)^*\Big) \ .$$

Multiplying both sides by 1_{B_n} we obtain:

$$\begin{aligned} &\Big(\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+\Big)^* = \Big(\big(\sup_{\lambda \in \Lambda} M_n(f_\lambda)\big)^+\Big)^* \\ &= \Big(\big(\sup_{\lambda \in \Lambda} M_n(f_\lambda)\big)^*\Big)^+ = \Big(\sup_{\lambda \in \Lambda} M_n(f_\lambda)\Big)^* \cdot 1_{B_n} \\ &\leq \Big(\sup_{\lambda \in \Lambda} f_\lambda\Big)^* \cdot 1_{B_n} + T\Big(\big(\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+\big)^*\Big) \cdot 1_{B_n} \leq \end{aligned}$$

$$\leq \Big(\sup_{\lambda \in \Lambda} f_\lambda \Big)^* \cdot 1_{B_n} + \; T\Big(\big(\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+ \big)^* \Big) \ .$$

Integrating both sides we get:

$$\int^* \sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+ \; d\mu$$

$$\leq \int_{B_n}^* \sup_{\lambda \in \Lambda} f_\lambda \; d\mu \; + \; \int T((\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+)^*) \; d\mu \ .$$

Finally, as above, this can be rewritten as follows:

$$\int^* \sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+ \, d\mu - \int T((\sup_{\lambda \in \Lambda} (M_n(f_\lambda))^+)^*) \; d\mu \leq \int_{B_n}^* \sup_{\lambda \in \Lambda} f_\lambda \; d\mu$$

and the proof of (2.86) follows from the contractibility of T .

Next suppose that there is no $g \in L^1(\mu)$ satisfying $\sup_{\lambda \in \Lambda} f_\lambda \leq g$. In this case we have $\int^* \sup_{\lambda \in \Lambda} f_\lambda \; d\mu = +\infty$. Let $n \geq 1$ be given and fixed. Then by subadditivity we get:

$$\int^* \sup_{\lambda \in \Lambda} f_\lambda \; d\mu \;\leq\; \int_{C_n}^* \sup_{\lambda \in \Lambda} f_\lambda \; d\mu \; + \; \int_{C_n^c}^* \sup_{\lambda \in \Lambda} f_\lambda \; d\mu$$

with C_n being equal either A_n or B_n . However, in either of the cases, on the set C_n^c we evidently have $\sup_{\lambda \in \Lambda} f_\lambda \leq \sup_{\lambda \in \Lambda} M_n(f_\lambda) \leq 0$. Therefore $\int_{C_n}^* \sup_{\lambda \in \Lambda} f_\lambda \; d\mu = +\infty$, and the proof of theorem is complete. □

Corollary 2.15 (The uniform ergodic inequality 1993)

Under the hypotheses of Theorem 2.14 suppose moreover that μ *is finite, and that* $T(\mathbf{1}) = \mathbf{1}$ *. Let us denote* $A_{n,t} = \{ \sup_{\lambda \in \Lambda} R_n(f_\lambda) > t \}$ *and* $B_{n,t} = \{ (\sup_{\lambda \in \Lambda} R_n(f_\lambda))^* > t \}$ *for* $n \geq 1$ *and* $t > 0$ *. Then we have:*

$$\int_{A_{n,t}}^* \sup_{\lambda \in \Lambda} (f_\lambda - t) \; d\mu \;\geq\; 0 \tag{2.90}$$

$$\mu_* \Big\{ \sup_{\lambda \in \Lambda} R_n(f_\lambda) > t \Big\} \;\leq\; \frac{1}{t} \int_{A_{n,t}}^* \sup_{\lambda \in \Lambda} \, f_\lambda \; d\mu \tag{2.91}$$

$$\int_{B_{n,t}}^* \sup_{\lambda \in \Lambda} (f_\lambda - t) \; d\mu \;\geq\; 0 \tag{2.92}$$

$$\mu^* \Big\{ \sup_{\lambda \in \Lambda} R_n(f_\lambda) > t \Big\} \;\leq\; \frac{1}{t} \int_{B_{n,t}}^* \sup_{\lambda \in \Lambda} \, f_\lambda \; d\mu \tag{2.93}$$

for all $n \geq 1$ *and all* $t > 0$ *.*

Proof. Let $n \geq 1$ and $t > 0$ be given and fixed. Consider $g_\lambda = f_\lambda - t$ for $\lambda \in \Lambda$. Since $T(\mathbf{1}) = \mathbf{1}$, then it is easily verified that $A_{n,t} = \{ \sup_{\lambda \in \Lambda} M_n(g_\lambda) > 0 \}$ and therefore $B_{n,t} = \{ (\sup_{\lambda \in \Lambda} M_n(g_\lambda))^* > 0 \}$. Hence by subadditivity and Theorem 2.14 we obtain:

$$\int_{C_{n,t}}^* \sup_{\lambda \in \Lambda} f_\lambda \; d\mu \; + \; \int_{C_{n,t}}^* (-t) \; d\mu \;\geq\; \int_{C_{n,t}}^* \sup_{\lambda \in \Lambda} g_\lambda \; d\mu \;\geq\; 0$$

with $C_{n,t}$ being equal either $A_{n,t}$ or $B_{n,t}$. Thus the proof follows straightforwardly from the facts that $\int^*_{A_{n,t}} (-t)\ d\mu = -t \cdot \mu_*(A_{n,t})$ and $\int^*_{B_{n,t}} (-t)\ d\mu = -t \cdot \mu(B_{n,t})$.

$\square$

Remark 2.16

Under the hypotheses of Theorem 2.14 and Corollary 2.15 respectively, suppose that $(\Lambda, \mathcal{B})$ is an *analytic space* (see [43] p.12). We point out that any *polish space* is an analytic metric space. Suppose moreover that the map $(x, \lambda) \mapsto f_\lambda(x)$ from $X \times \Lambda$ into $\bar{\mathbf{R}}$ is measurable with respect to the product σ-algebra $\mathcal{A} \times \mathcal{B}$ and Borel σ-algebra $\mathcal{B}(\bar{\mathbf{R}})$. Then by the *projection theorem* (see [43] p.13-14) we may conclude that the map $x \mapsto \sup_{\lambda \in \Lambda} f_\lambda(x)$ is *universally measurable* from X into $\bar{\mathbf{R}}$, which means μ-measurable with respect to any measure μ on $\mathcal{A}$. In this way all upper μ-integrals, outer and inner μ-measures in Theorem 2.14 and Corollary 2.15 become the ordinary ones (without stars). Moreover, then both Theorem 2.14 and Corollary 2.15 extend to the case where the supremum in the definitions of functions M and R is taken over all integers. We leave the precise formulation of these facts and remaining details to the reader.

2.4 The uniform ergodic theorem for dynamical systems

1. The main object under consideration in this section is the given *ergodic* dynamical system $(X, \mathcal{A}, \mu, T)$. Thus $(X, \mathcal{A}, \mu)$ is a probability space, and T is an ergodic measure-preserving transformation of X (see Paragraph 4 in Section 1.1). We further assume that a parameterized family $\mathcal{F} = \{ f_\lambda \mid \lambda \in \Lambda \}$ of measurable maps from X into $\mathbf{R}$ is given, such that:

(2.94) $$\int^* \sup_{\lambda \in \Lambda} |f_\lambda|\ d\mu < \infty\ .$$

In particular, we see that each f_λ belongs to $L^1(\mu)$, and therefore for all $\lambda \in \Lambda$ by Birkhoff's Ergodic Theorem 1.6 we have:

(2.95) $$\frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\lambda) \to M(\lambda) \quad \mu\text{-a.s.}$$

as $n \to \infty$. The limit M is called the *μ-mean function* of $\mathcal{F}$, and we have $M(\lambda) = \int f_\lambda\, d\mu$ for all $\lambda \in \Lambda$. Hence we see that M belongs to $B(\Lambda)$ whenever (2.94) is satisfied, where $B(\Lambda)$ denotes the set of all bounded real valued functions on Λ .

2. The purpose of this section is to present a solution to the following problem. Determine conditions which are necessary and sufficient for the uniform convergence to be valid:

(2.96) $$\sup_{\lambda \in \Lambda} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big| \to 0$$

as $n \to \infty$. More precisely, we shall consider three convergence concepts in (2.96), and these are the $(a.s.)^*$-convergence, the $(L^1)^*$-convergence, and the (μ^*)-convergence. Given a sequence of arbitrary maps $\{ Z_n \mid n \geq 1 \}$ from X into $\mathbf{R}$, we recall that $Z_n \to 0\ (a.s.)^*$ if $|Z_n|^* \to 0$ μ-a.s., that $Z_n \to 0\ (L^1)^*$ if $|Z_n|^* \to 0$ in μ-mean, and that $Z_n \to 0\ (\mu^*)$ if $|Z_n|^* \to 0$ in μ-measure. For more information in this direction see Paragraph 4 in Section 2.2 and recall (2.22). It should be noted that if all the maps under consideration are measurable, then the convergence concepts stated above coincide with the usual concepts of μ-a.s. convergence, convergence in

μ-mean, and convergence in μ-measure, respectively. This is true in quite a general setting as described in Remark 2.16 above.

In order to handle measurability problems appearing in the sequel, we shall often assume that the transformation T is μ-perfect. This means that for every $A \in \mathcal{A}$ there exists $B \in \mathcal{A}$ satisfying $B \subset T(A)$ and $\mu(A \backslash T^{-1}(B)) = 0$. We can require equivalently that $\mu^*(T^{-1}(C)) = 1$ whenever C is a subset of X satisfying $\mu^*(C) = 1$. Under this assumption we have:

(2.97) $$(g \circ T)^* = g^* \circ T$$

(2.98) $$\int^* g \circ T \, d\mu = \int^* g \, d\mu$$

whenever $g : X \to \bar{\mathbf{R}}$ is an arbitrary map. For more details see [62].

The dynamical system $(X, \mathcal{A}, \mu, T)$ is said to be *perfect*, if T is μ-perfect. The best known sufficient condition for the μ-perfectness of T is the following:

(2.99) $$T(\mathcal{A}) \subset \mathcal{A}_\mu$$

where $\mathcal{A}_\mu$ denotes the μ-completion of $\mathcal{A}$. This condition is by the image theorem satisfied whenever $(X, \mathcal{A})$ is an analytic metric space (see [43] p.13). Moreover, if $(X, \mathcal{A})$ is the countable product $(S^{\mathbf{N}}, \mathcal{B}^{\mathbf{N}})$ of copies of a measurable space $(S, \mathcal{B})$, then by the projection theorem we see that condition (2.99) is satisfied whenever $(S, \mathcal{B})$ is an analytic space (see [43] p.13). We remind again that any polish space is an analytic metric space.

3. We return to the problem (2.96) by reminding that its meaning and applications are described in Section 1.3 and Section 1.4. Motivated by the results from Section 2.2 we shall say that the family $\mathcal{F} = \{ f_\lambda \mid \lambda \in \Lambda \}$ is *eventually totally bounded in μ-mean* with respect to T , if the following condition is satisfied:

(2.100) For every $\varepsilon > 0$ there exists $\gamma_\varepsilon \in \Gamma(\Lambda)$ such that:

$$\inf_{n \geq 1} \frac{1}{n} \int^* \sup_{\lambda', \lambda'' \in A} \Big| \sum_{j=0}^{n-1} T^j (f_{\lambda'} - f_{\lambda''}) \Big| \, d\mu < \varepsilon$$

for all $A \in \gamma_\varepsilon$.

This definition should be compared with definition (2.29). Here and in the sequel $\Gamma(\Lambda)$ denotes the family of all finite coverings of Λ . We recall that a finite covering of Λ is any family $\gamma = \{ A_1, \ldots, A_n \}$ of non-empty subsets of Λ satisfying $\Lambda = \cup_{j=1}^n A_j$ with $n \geq 1$.

4. Our next aim is to show that under (2.94), the condition (2.100) is equivalent to the uniform convergence in (2.96), with respect to any of the convergence concepts stated above. It should be noted that our method below in essence relies upon the uniform ergodic lemma (Theorem 2.14) from the previous section. The first result is just a reformulation of Theorem 2.4 and is included for the sake of completeness.

Theorem 2.17

Let $(X, \mathcal{A}, \mu, T)$ be a perfect ergodic dynamical system, and let $\mathcal{F} = \{ f_\lambda \mid \lambda \in \Lambda \}$ be a parameterized family of measurable maps from X into $\mathbf{R}$ satisfying (2.94) above. If $\mathcal{F}$ is *eventually totally bounded in μ-mean with respect to T, then we have:*

$$\sup_{\lambda \in \Lambda} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big| \to 0 \quad (a.s.)^* \ \& \ (L^1)^* \tag{2.101}$$

as $n \to \infty$, where M is the μ-mean function of $\mathcal{F}$.

Proof. Let $\varepsilon > 0$ be given and fixed, then by our assumption there exists $\gamma_\varepsilon \in \Gamma(\Lambda)$ such that:

$$\inf_{n \geq 1} \int^* \sup_{\lambda', \lambda'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \, d\mu < \varepsilon \tag{2.102}$$

for all $A \in \gamma_\varepsilon$. Since $M \in B(\Lambda)$, there is no restriction to assume that we also have:

$$\sup_{\lambda', \lambda'' \in A} |M(\lambda') - M(\lambda'')| < \varepsilon \tag{2.103}$$

for all $A \in \gamma_\varepsilon$. Choosing a point $t_A \in A$ for every $A \in \gamma_\varepsilon$, from (2.103) we get:

$$\sup_{\lambda \in \Lambda} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big| = \max_{A \in \gamma_\varepsilon} \sup_{\lambda \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big|$$

$$\leq \max_{A \in \gamma_\varepsilon} \Big(\sup_{\lambda', \lambda'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| + \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda_A}) - M(\lambda_A) \Big|$$

$$+ \sup_{\lambda', \lambda'' \in A} |M(\lambda') - M(\lambda'')| \Big) \leq \max_{A \in \gamma_\varepsilon} \sup_{\lambda', \lambda'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big|$$

$$+ \max_{A \in \gamma_\varepsilon} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda_A}) - M(\lambda_A) \Big| + \varepsilon$$

for all $n \geq 1$. Hence by (2.95) we easily obtain:

$$\limsup_{n \to \infty} \Big(\sup_{\lambda \in \Lambda} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big| \Big)^*$$

$$\leq \max_{A \in \gamma_\varepsilon} \limsup_{n \to \infty} \Big(\sup_{\lambda', \lambda'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \Big)^* + \varepsilon$$

From this inequality and (2.102) we see that the $(a.s.)^*$-convergence in (2.101) will be established as soon as we deduce the inequality:

$$\limsup_{n \to \infty} \Big(\sup_{\lambda', \lambda'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \Big)^* \leq \inf_{n \geq 1} \int^* \sup_{\lambda', \lambda'' \in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \, d\mu \tag{2.104}$$

for all $A \subset \Lambda$. We leave this inequality to be established with some facts of independent interest in the next proposition, and proceed with the $(L^1)^*$-convergence in (2.101). From (2.97) we get:

$$\Big(\sup_{\lambda\in\Lambda} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big| \Big)^* \leq \frac{1}{n} \sum_{j=0}^{n-1} \Big(\sup_{\lambda\in\Lambda} \big| T^j(f_\lambda) \big| \Big)^* + \sup_{\lambda\in\Lambda} |M(\lambda)|$$
$$= \frac{1}{n} \sum_{j=0}^{n-1} T^j \Big((\sup_{\lambda\in\Lambda} |f_\lambda|)^* \Big) + \sup_{\lambda\in\Lambda} |M(\lambda)|$$

for all $n \geq 1$. Hence from (2.94) and (2.9)+(2.13) it follows that the sequence on the left-hand side is uniformly integrable. Thus the $(L^1)^*$-convergence in (2.101) follows from the $(a.s.)^*$-convergence, and the proof is complete. □

Proposition 2.18

Under the hypotheses of Theorem 2.17 we have:

(2.105) $$\limsup_{n\to\infty} \Big(\sup_{\lambda',\lambda''\in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \Big)^* = C \quad \mu\text{-a.s.}$$

(2.106) $$C \leq \inf_{n\geq 1} \int^* \sup_{\lambda',\lambda''\in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \, d\mu$$

(2.107) $$\inf_{n\geq 1} \int^* \sup_{\lambda',\lambda''\in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \, d\mu = \limsup_{n\to\infty} \int^* \sup_{\lambda',\lambda''\in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \, d\mu$$

where A is an arbitrary subset of Λ , and C is a real number depending on A .

Proof. Let $A \subset \Lambda$ be given and fixed. Let us denote:

$$Z_n(x) = \sup_{\lambda',\lambda''\in A} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''})(x) \Big|$$

for all $x \in X$ and all $n \geq 1$. Then the proof can be carried out respectively as follows.

(2.105): Since T is ergodic, it is enough to show that:

(2.108) $$\limsup_{n\to\infty} Z_n^* \circ T = \limsup_{n\to\infty} Z_n^* \quad \mu\text{-a.s.}$$

For this it should be noted that we have:

$$Z_n \circ T = \sup\nolimits_{\lambda',\lambda''\in A} \Big| \tfrac{n+1}{n} \cdot \tfrac{1}{n+1} \sum\nolimits_{j=0}^{n} T^j(f_{\lambda'} - f_{\lambda''}) - \tfrac{1}{n} (f_{\lambda'} - f_{\lambda''}) \Big|$$

for all $n \geq 1$. Hence we easily find:

$$\tfrac{n+1}{n} Z_{n+1} - \tfrac{2}{n} \sup\nolimits_{\lambda\in\Lambda} |f_\lambda| \leq Z_n \circ T \leq \tfrac{n+1}{n} Z_{n+1} + \tfrac{2}{n} \sup\nolimits_{\lambda\in\Lambda} |f_\lambda|$$

for all $n \geq 1$. Taking the upper μ-envelopes on both sides we obtain:

$$\tfrac{n+1}{n} Z_{n+1}^* - \tfrac{2}{n} \big(\sup\nolimits_{\lambda\in\Lambda} |f_\lambda| \big)^* \leq \big(Z_n \circ T \big)^* \leq \tfrac{n+1}{n} Z_{n+1}^* + \tfrac{2}{n} \big(\sup\nolimits_{\lambda\in\Lambda} |f_\lambda| \big)^*$$

for all $n \geq 1$. Letting $n \to \infty$ and using (2.94)+(2.97), we obtain (2.108), and the proof of

(2.105) is complete.

(2.106): Here we essentially use the uniform ergodic lemma (Theorem 2.14) from the last section. For this we denote $S_n(f) = \sum_{j=0}^{n-1} T^j f$ and $M_n(f) = \max_{1 \leq j \leq n} S_j(f)$, whenever $f \in L^1(\mu)$ and $n \geq 1$. Let us for fixed $N, m \geq 1$ and $\varepsilon > 0$ consider the following set:

$$B_{N,m,\varepsilon} = \Big\{ \Big(\sup_{\lambda',\lambda'' \in A} M_m\big(|S_N(f_{\lambda'} - f_{\lambda''})/N| - (C-\varepsilon) \big) \Big)^* > 0 \Big\} .$$

Then by Theorem 2.14 we may conclude:

$$\int_{B_{N,m,\varepsilon}}^* \sup_{\lambda',\lambda'' \in A} \big(|S_N(f_{\lambda'} - f_{\lambda''})/N| - (C-\varepsilon) \big) \, d\mu \geq 0 \ .$$

Hence by subadditivity of the upper μ-integral we obtain:

$$\int_{B_{N,m,\varepsilon}}^* \sup_{\lambda',\lambda'' \in A} |S_N(f_{\lambda'} - f_{\lambda''})/N| \, d\mu \ \geq \ (C-\varepsilon) \, \mu(B_{N,m,\varepsilon}) \ .$$

Therefore by the monotone convergence theorem for upper integrals (see [62]), in order to complete the proof of (2.106), it is enough to show that $\mu(B_{N,m,\varepsilon}) \uparrow 1$ as $m \to \infty$, with $N \geq 1$ and $\varepsilon > 0$ being fixed. We shall establish this fact by proving the following inequality:

(2.109) $$\Big(\sup_{\lambda',\lambda'' \in A} |S_n(f_{\lambda'} - f_{\lambda''})/n| \Big)^* \leq \Big(\sup_{\lambda',\lambda'' \in A} \sup_{m \geq 1} S_m\big(|S_N(f_{\lambda'} - f_{\lambda''})/N| \big) \Big)^* + \Big(\tfrac{1}{n} \sup_{\lambda',\lambda'' \in A} \big| \textstyle\sum_{N[n/N] \leq j < n} T^j (f_{\lambda'} - f_{\lambda''}) \big| \Big)^*$$

for all $n \geq N$. For this it should be noted that we have:

(2.110) $$|S_n(f_{\lambda'} - f_{\lambda''})/n| \leq |S_{N[n/N]}(f_{\lambda'} - f_{\lambda''})/n| + \tfrac{1}{n} \big| \textstyle\sum_{N[n/N] \leq j < n} T^j (f_{\lambda'} - f_{\lambda''}) \big|$$

for all $n \geq N$ and all $\lambda', \lambda'' \in A$. Moreover, given $n \geq N$, taking $m \geq 1$ large enough we get:

$$\begin{aligned} S_m\big(|S_N(f_{\lambda'} - f_{\lambda''})| \big) \ &= \ \big| \textstyle\sum_{j=0}^{N-1} T^j (f_{\lambda'} - f_{\lambda''}) \big| \ + \ \big| \textstyle\sum_{j=1}^{N} T^j (f_{\lambda'} - f_{\lambda''}) \big| \ + \ \dots \\ &+ \big| \textstyle\sum_{j=N}^{2N-1} T^j (f_{\lambda'} - f_{\lambda''}) \big| \ + \ \dots \\ &\geq \big| \textstyle\sum_{j=0}^{N-1} T^j (f_{\lambda'} - f_{\lambda''}) \big| + \big| \textstyle\sum_{j=N}^{2N-1} T^j (f_{\lambda'} - f_{\lambda''}) \big| + \dots + \big| \textstyle\sum_{j=N([n/N]-1)}^{N[n/N]-1} T^j (f_{\lambda'} - f_{\lambda''}) \big| \\ &\geq \ |S_{N[n/N]}(f_{\lambda'} - f_{\lambda''})| \end{aligned}$$

for all $\lambda', \lambda'' \in A$. Hence we obtain:

(2.111) $$|S_{N[n/N]}(f_{\lambda'} - f_{\lambda''})/n| \leq S_m\big(|S_N(f_{\lambda'} - f_{\lambda''})/n| \big) \leq S_m\big(|S_N(f_{\lambda'} - f_{\lambda''})/N| \big)$$

with $n, N, m \geq 1$ as above. Thus (2.109) follows straightforwardly by (2.110) and (2.111). In addition, for the last term in (2.109) we have:

$$\Big(\tfrac{1}{n} \ \sup_{\lambda',\lambda''\in A} \big| \textstyle\sum_{N[n/N]\le j<n} T^j(f_{\lambda'} - f_{\lambda''}) \ \big| \Big)^*$$
$$\le \ \tfrac{2}{n} \textstyle\sum_{N[n/N]\le j<n} T^j\big(\ (\sup_{\lambda\in A} |f_\lambda|)^*\big)$$
$$= \ 2 \left(\tfrac{1}{n} S_n\big(\ (\sup_{\lambda\in A} |f_\lambda|)^*\big) - \tfrac{N[n/N]}{n} \cdot \tfrac{1}{N[n/N]} \, S_{N[n/N]}\big(\ (\sup_{\lambda\in A} |f_\lambda|)^*\big) \right)$$

for all $n \ge N$. Since $N[n/N]/n \to 1$ as $n \to \infty$, then by (2.94) and Birkhoff's Ergodic Theorem 1.6, the right-hand side tends to zero μ-a.s. as $n \to \infty$. Letting $n \to \infty$ in (2.109) therefore we obtain:

$$\text{(2.112)} \qquad \limsup_{n\to\infty} \Big(\sup_{\lambda',\lambda''\in A} |S_n(f_{\lambda'} - f_{\lambda''})/n| \Big)^* \le \Big(\sup_{\lambda',\lambda''\in A} \ \sup_{m\ge 1} \ S_m\big(\ |S_N(f_{\lambda'} - f_{\lambda''})/N| \ \big) \Big)^*$$

for all $N \ge 1$. Finally, from (2.105) and (2.112) we get:

$$\lim_{m\to\infty} \mu(B_{N,m,\varepsilon}) = \lim_{m\to\infty} \mu^* \Big\{ \sup_{\lambda',\lambda''\in A} M_m\big(\ |S_N(f_{\lambda'} - f_{\lambda''})/N| \ \big) > (C - \varepsilon) \Big\}$$
$$= \ \mu^* \Big\{ \sup_{\lambda',\lambda''\in A} \ \sup_{m\ge 1} \ S_m\big(\ |S_N(f_{\lambda'} - f_{\lambda''})/N| \ \big) > (C - \varepsilon) \Big\}$$
$$= \ \mu \Big\{ \Big(\sup_{\lambda',\lambda''\in A} \ \sup_{m\ge 1} \ S_m\big(\ |S_N(f_{\lambda'} - f_{\lambda''})/N| \ \big) \Big)^* > (C - \varepsilon) \Big\}$$
$$\ge \ \mu \Big\{ \limsup_{n\to\infty} \Big(\sup_{\lambda',\lambda''\in A} |S_n(f_{\lambda'} - f_{\lambda''})/n| \Big)^* > (C - \varepsilon) \Big\} = 1 \ .$$

This fact completes the proof of (2.106).

(2.107): It should be noted that we have:

$$\Big(\sup_{\lambda',\lambda''\in A} \ \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \Big)^* \le \ \frac{2}{n} \ \sum_{j=0}^{n-1} T^j\big(\ (\sup_{\lambda\in A} |f_\lambda|)^*\big)$$

for all $n \ge 1$. Therefore by (2.94) and (2.9)+(2.13) it follows that the sequence on the left-hand side is uniformly integrable. Thus by Fatou's lemma we can conclude:

$$\limsup_{n\to\infty} \int^* \sup_{\lambda',\lambda''\in A} \ \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big| \ d\mu \ \le C \ .$$

Hence (2.107) follows straightforwardly by (2.106). This fact completes the proof. □

Remark 2.19

Under the hypotheses of Theorem 2.17 and Proposition 2.18 it is easily verified that in the case when the map $x \mapsto \sup_{\lambda',\lambda''\in A} \big| \sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''})(x) \big|$ is μ-measurable as a map from X into $\mathbf{R}$ for all $n \ge 1$ and given $A \subset \Lambda$, the assumption of μ-perfectness of T is not needed for their conclusions remain valid.

Remark 2.20

Besides the given proof, there are at least two alternative proofs of the result in Proposition 2.18. First, we may note that if T is μ-perfect, then T^{j-1} is μ-perfect for all $j \ge 1$. This fact easily implies that the map $x \mapsto (T^0(x), T^1(x), \ \dots \ )$ is μ-perfect as a map from $(X, \mathcal{A})$ into

$(X^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$. Hence we see that if T is assumed to be μ-perfect, then the hypotheses of Theorem 2.4 and Proposition 2.5 are satisfied with $\xi_j = T^{j-1}$ for $j \geq 1$. The proof then could be carried out in exactly the same way as the proof of Proposition 2.5.

Second, we shall note that the result of Proposition 2.18 could also be deduced by Kingman's Subadditive Ergodic Theorem 1.8. Indeed, let us in the notation of Proposition 2.18 denote:

$$g_n = \Big(\sup_{\lambda', \lambda'' \in A} \Big| \sum_{j=0}^{n-1} T^j (f_{\lambda'} - f_{\lambda''}) \Big| \Big)^* \tag{2.113}$$

for all $n \geq 1$, and let us assume that T is μ-perfect. Then it is quite easily verified that the sequence $\{ g_n \mid n \geq 1 \}$ is T-subadditive in $L^1(\mu)$. Therefore (2.105)-(2.107) in Proposition 2.18 follow from Kingman's Theorem 1.8. Moreover, although irrelevant for our purposes, it should be noted that in this way we immediately obtain a more precise information: We have the equalities in (2.105) and (2.106) with the limit in (2.105) instead of the limit superior.

In this context it may also be instructive to recall the classical fact that for a subadditive sequence of real numbers $\{ \gamma_n \mid n \geq 1 \}$ we have $\lim_{n \to \infty} \gamma_n / n = \inf_{n \geq 1} \gamma_n / n$. We shall conclude this remark by expressing our belief that the method of proof presented in Proposition 2.18 above, relying upon the uniform ergodic lemma (Theorem 2.14) from the previous section, may be useful in the investigation of similar problems in more general operator cases as well.

Remark 2.21

It should be noted that if we require the infimum in (2.100) to be attained for $n = 1$, then we obtain the Blum-DeHardt condition of Theorem 2.1, which is the best known sufficient condition for the uniform law of large numbers (at least in the independent case). There is an example (presented in Section 2.5.1 below) showing that a parameterized family could be eventually totally bounded in the mean (with respect to a measure-preserving transformation) without satisfying the Blum-DeHardt condition. This in fact is not surprising since we show in the next theorem that the property of eventually totally bounded in the mean characterizes the uniform ergodic theorem for dynamical systems, and thus the uniform laws of large numbers as well (see Section 1.3 and Corollary 2.8). We may thus conclude that the condition (2.100) offers a characterization of the uniform ergodic theorem for dynamical systems which contains the best known sufficient condition as a particular case. In this context it is interesting to recall that the Blum-DeHardt law of large numbers contains Mourier's classic law of large numbers (see Corollary 2.3).

In order to state the next theorem we shall once again recall that $B(\Lambda)$ denotes the Banach space of all bounded real valued functions on Λ with respect to the sup-norm. We denote by $C(B(\Lambda))$ the set of all bounded continuous functions from $B(\Lambda)$ into $\mathbf{R}$, and by $\mathcal{K}(B(\Lambda))$ we denote the family of all compact subsets of $B(\Lambda)$. A pseudo-metric d on a set Λ is said to be *totally bounded*, if Λ can be covered by finitely many d-balls of any given radius $\varepsilon > 0$. A pseudo-metric d on a set Λ is said to be an *ultra pseudo-metric*, if it satisfies $d(\lambda_1, \lambda_2) \leq d(\lambda_1, \lambda_3) \vee d(\lambda_3, \lambda_2)$ whenever $\lambda_1, \lambda_2, \lambda_3 \in \Lambda$.

Theorem 2.22

Let $(X, \mathcal{A}, \mu, T)$ *be a perfect ergodic dynamical system, let* $\mathcal{F} = \{ f_\lambda \mid \lambda \in \Lambda \}$ *be a parameterized family of measurable maps from* X *into* $\mathbf{R}$ *satisfying (2.94) above, and let* M *be* *the μ-mean function of $\mathcal{F}$. Then the following eight statements are equivalent:*

(2.114) *The family $\mathcal{F}$ is eventually totally bounded in μ-mean with respect to T*

(2.115) $$\sup_{\lambda\in\Lambda} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big| \to 0 \ \ (a.s.)^*$$

(2.116) $$\sup_{\lambda\in\Lambda} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big| \to 0 \ \ (L^1)^*$$

(2.117) $$\sup_{\lambda\in\Lambda} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\lambda) - M(\lambda) \Big| \to 0 \ \ (\mu^*)$$

(2.118) $\frac{1}{n}\sum_{j=0}^{n-1} T^j(f) \to M$ *weakly in* $B(\Lambda)$ *, that is:*

$$\lim_{n\to\infty} \int^* F\Big(\frac{1}{n}\sum_{j=0}^{n-1} T^j(f)\Big)\, d\mu = F(M)$$

for all $F \in C(B(\Lambda))$

(2.119) *The sequence* $\Big\{ \frac{1}{n}\sum_{j=0}^{n-1} T^j(f) \mid n \geq 1 \Big\}$ *is eventually tight in* $B(\Lambda)$ *, that is:*

$$\limsup_{n\to\infty} \int^* F\Big(\frac{1}{n}\sum_{j=0}^{n-1} T^j(f)\Big)\, d\mu \leq \varepsilon$$

for some $K_\varepsilon \in \mathcal{K}(B(\Lambda))$ *, and for all* $F \in C(B(\Lambda))$ *satisfying* $0 \leq F \leq 1_{B(\Lambda)\setminus K_\varepsilon}$ *, whenever* $\varepsilon > 0$

(2.120) *There exists a totally bounded ultra pseudo-metric d on Λ such that the condition is satisfied:*

$$\lim_{r\downarrow 0} \inf_{n\geq 1} \int^* \sup_{d(\lambda',\lambda'')<r} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_{\lambda'} - f_{\lambda''}) \Big|\, d\mu = 0$$

(2.121) *For every $\varepsilon > 0$ there exist a totally bounded pseudo-metric d_ε on Λ and $r_\varepsilon > 0$ such that:*

$$\inf_{n\geq 1} \mu^*\Big\{ \sup_{d_\varepsilon(\lambda,\lambda')<\varepsilon} \Big| \frac{1}{n}\sum_{j=0}^{n-1} T^j(f_\lambda - f_{\lambda'}) \Big| > \varepsilon \Big\} < \varepsilon$$

for all $\lambda \in \Lambda$.

Proof. Since T is μ-perfect, then T^{j-1} is μ-perfect for all $j \geq 1$. This fact easily implies that the map $x \mapsto (T^0(x), T^1(x), \dots)$ is μ-perfect as a map from $(X, \mathcal{A})$ into $(X^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$. Putting $\xi_j = T^{j-1}$ for $j \geq 1$ we see that the hypotheses of Theorem 2.7, Corollary 2.8, Theorem 2.10, Corollary 2.11 and Theorem 2.13 are satisfied. The result therefore follows from Corollary 2.8, Theorem 2.10, Corollary 2.11 and Theorem 2.13. These facts complete the proof.

□

2.5 Examples and complements

This section contains examples and complementary facts which are aimed to support and clarify the results from the previous four sections of this chapter.

2.5.1 Dudley's counter-example. The example is from [18] and shows that the Blum-DeHardt sufficient condition from Theorem 2.1 is generally not necessary for the uniform law of large numbers. In other words, if the infimum in (2.29) and (2.100) is not attained for $n = 1$, it does not mean that the uniform ergodic theorem fails (see Theorem 2.4 and Theorem 2.17). In the construction below we need the following lemma (see [18] p.14-16).

Lemma 2.23 (Chernoff 1952)

Let $0 \leq p = 1 - q \leq 1$ *and* $n \geq 1$ *be given. Then the estimates are valid:*

$$\sum_{j=k}^{n} \binom{n}{j} p^j q^{n-j} \leq (np/k)^k (nq/(n-k))^{n-k} \tag{2.122}$$

$$\sum_{j=k}^{n} \binom{n}{j} p^j q^{n-j} \leq (np/k)^k \exp(k-np) \tag{2.123}$$

for all $k \geq np$.

Proof. Let the random variable $X \sim B(n,p)$ be from the binomial distribution. Then by Markov's inequality we get:

$$\sum_{j=k}^{n} \binom{n}{j} p^j q^{n-j} = P\{X \geq k\} = P\{\exp(\xi X) \geq \exp(\xi k)\} \leq \exp(-\xi k)\, E\big(\exp(\xi X)\big) =$$
$$= \exp(-\xi k) \sum_{j=0}^{n} \exp(\xi j) \binom{n}{j} p^j q^{n-j} = \exp(-\xi k)\, (pe^{\xi} + q)^n$$

for all $\xi \geq 0$. Putting $\xi = \log(kq/(n-k)p)$ in the last expression, we obtain (2.122). Inserting $x \leq \exp(x-1)$ in (2.122) with $x = nq/(n-k)$, we obtain (2.123). This completes the proof. □

Example 2.24 (Dudley 1982)

Let $\{\xi_j\}_{j\geq 1}$ be a sequence of independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in the measurable space $(S, \mathcal{A})$ and common distribution law π . Let $\{C_m\}_{m\geq 1}$ be a sequence of independent sets in $(S, \mathcal{A}, \pi)$ with $\pi(C_m) = 1/m$ for all $m \geq 1$. Then we have:

$$\sup_{m\geq 1} \Big| \frac{1}{n} \sum_{j=1}^{n} 1_{C_m}(\xi_j) - \pi(C_m) \Big| \to 0 \quad P\text{-a.s.} \tag{2.124}$$

as $n \to \infty$.

For this, let $\varepsilon > 0$ be given and fixed. Put $m_0 = [e/\varepsilon]$, then by (3.3) below (with Corollary 2.8) it is enough to show:

$$\hat{P}\Big\{ \sup_{m\geq m_0} \Big| \frac{1}{n} \sum_{j=1}^{n} 1_{C_m}(\xi_j) \cdot \varepsilon_j \Big| > \varepsilon \Big\} \to 0 \tag{2.125}$$

as $n \to \infty$. Here $\hat{P} = P \otimes P_\varepsilon$ is the product probability, and $\{\varepsilon_j\}_{j\geq 1}$ is a Rademacher sequence defined on the probability space $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon)$, which is understood to be independent from $\{\xi_j\}_{j\geq 1}$.

For (2.125) we first note that $e/m\varepsilon < 1$ for all $m > m_0$, and thus $(e/m\varepsilon)^{[n\varepsilon]+1} \leq (e/m\varepsilon)^2$ for all $n \geq 1/\varepsilon$. Using this fact, the subgaussian inequality (3.1), and (2.123), we obtain:

$$\hat{P}\Big\{ \sup_{m \geq m_0} \Big| \frac{1}{n} \sum_{j=1}^{n} 1_{C_m}(\xi_j) \cdot \varepsilon_j \Big| > \varepsilon \Big\} \leq \sum_{m > m_0} E\Big(P_\varepsilon \Big\{ \Big| \sum_{j=1}^{n} 1_{C_m}(\xi_j) \cdot \varepsilon_j \Big| > n\varepsilon \Big\}\Big)$$

$$\leq \sum_{m > m_0} E\bigg(2 \exp\big(-\varepsilon^2 n^2 / 2 \textstyle\sum_{j=1}^{n} 1_{C_m}(\xi_j) \big) \cdot 1_{\{\sum_{j=1}^{n} 1_{C_m}(\xi_j) \,\geq\, [n\varepsilon]+1\}} \bigg)$$

$$= \sum_{m > m_0} 2 \sum_{j=[n\varepsilon]+1}^{n} \exp\big(-\varepsilon^2 n^2 / 2j \big) \tbinom{n}{j} (1/m)^j (1-1/m)^{n-j}$$

$$\leq 2 \exp\big(-\varepsilon^2 n / 2 \big) \sum_{m > m_0} \big(n/m([n\varepsilon]+1) \big)^{[n\varepsilon]+1} \exp\big([n\varepsilon]+1-n/m \big)$$

$$\leq 2 \exp\big(-\varepsilon^2 n / 2 \big) \sum_{m > m_0} \big(e/m\varepsilon \big)^{[n\varepsilon]+1} \exp(-n/m)$$

$$\leq 2 \exp\big(-\varepsilon^2 n / 2 \big) \frac{e^2}{\varepsilon^2} \sum_{m > m_0} \frac{1}{m^2} \leq \frac{\pi^2}{3} \frac{e^2}{\varepsilon^2} \exp\big(-\varepsilon^2 n / 2 \big)$$

for all $n \geq 1/\varepsilon$. Letting $n \to \infty$ we obtain (2.125), and the proof of (2.124) is complete.

In addition, let us put $\mathcal{C} = \{ 1_{C_m} \mid m \geq 1 \}$. Then we claim (in the notation of Section 2.1):

(2.126) $$N_1[\varepsilon, \mathcal{C}] = +\infty$$

for all $0 < \varepsilon < 1$. Thus, the Blum-DeHardt sufficient condition from Theorem 2.1 is not fulfilled.

Otherwise, there would exist $0 < \varepsilon < 1$ and $g_1 \leq h_1, \ldots, g_N \leq h_N$ in $L^1(\pi)$ such that:

(2.127) $$\mathcal{C} \subset \textstyle\bigcup_{k=1}^{N} [g_k, h_k]$$

(2.128) $$\max_{1 \leq k \leq N} \int_S (h_k - g_k) \, d\pi < \varepsilon \ .$$

From (2.127) and the fact $\sum_{m=1}^{\infty} \pi(C_m) = +\infty$, we see that there exists $1 \leq k \leq N$ such that for the subsequence $\{ 1_{C_{m_j}} \mid j \geq 1 \}$ of $\mathcal{C}$ satisfying $g_k \leq 1_{C_{m_j}} \leq h_k$ for all $j \geq 1$, we have $\sum_{j=1}^{\infty} \pi(C_{m_j}) = +\infty$. Let $A = \cap_{j=1}^{\infty} C_{m_j}$ and $B = \cup_{j=1}^{\infty} C_{m_j}$. Then $\pi(A) = 0$ and $\pi(B) = 1$, by Borel-Cantelli's lemma (see Section 1.2). Moreover, we have $g_k \leq 1_A$ and $1_B \leq h_k$. Thus $\int g_k \, d\pi \leq 0$ and $\int h_k \, d\pi \geq 1$. Hence $\int (h_k - g_k) \, d\pi \geq 1$, which contradicts (2.128). The proof of (2.126) is complete.

2.5.2 Necessity of integrability. The purpose of this section is to show that the integrability of the sup-norm is a necessary condition for the uniform law of large numbers to hold in the i.i.d. case. The first result we present appears in [39] and [84].

1. Let $(S, \mathcal{A}, \pi)$ be a probability space, and let $(\Omega, \mathcal{F}, P)$ be the countable product $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}}, \pi^{\mathbf{N}})$. Let $\xi_j\big((s_i)_{i \geq 1}\big) = s_j$ be the j-th projection of $S^{\mathbf{N}}$ onto S for $(s_i)_{i \geq 1} \in S^{\mathbf{N}}$ and $j \geq 1$. Then $\{\xi_j\}_{j \geq 1}$ is a sequence of independent and identically distributed random variables defined on $(\Omega, \mathcal{F}, P)$ with values in $(S, \mathcal{A})$ and common distribution law π . Let T

be a set, and let $f : S \times T \to \mathbf{R}$ be a given map. We denote $S_n(f) = \sum_{j=1}^n f(\xi_j)$ for all $n \geq 1$. In the sequel we adopt the notation from Section 2.2.

Theorem 2.25

Let $\{\xi_j\}_{j\geq 1}$ be a sequence of independent and identically distributed random variables defined on $(\Omega, \mathcal{F}, P)$ as above. Let us assume that:

(2.129) $$\big\| \tfrac{1}{n} S_n(f) - M \big\|_T \to 0 \;\; (a.s.)$$

as $n \to \infty$ for $M \in B(T)$. Then we have:

(2.130) $$\int^* \|f(s)\|_T \; \pi(ds) < \infty \; .$$

Proof. The proof of this result is based upon the following three facts:

(2.131) Let $(S, \mathcal{A}, \pi)$ be a probability space, and let C_n be a subset of S satisfying $\pi^*(C_n) = 1$ for all $n \geq 1$. Then $\big(\pi^{\mathbf{N}}\big)^*\big(\prod_{n=1}^\infty C_n\big) = 1$.

(2.132) Let $\{\eta_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables satisfying $\limsup_{n\to\infty} |\eta_n|/n < \infty$ P-a.s. Then $\eta_1 \in L^1(P)$.

(2.133) **(The 0-1 law for the upper envelopes)** Let $(S, \mathcal{A}, \pi)$ be a probability space, and let $f : S \to \mathbf{R}_+$ be a given map. Then we have:

$$\pi^*\big(\{f \geq \alpha\} \cup \{\beta f \geq f^*\}\big) = 1$$

whenever $\alpha > 0$ and $\beta > 1$.

The arguments leading to (2.131)-(2.133) may be presented as follows.

(2.131): Put $\mathcal{A}_n = \sigma(\mathcal{A} \cup \{C_n\})$ and define $\pi_n = tr^*(\pi, C_n)$ for all $n \geq 1$. In other words π_n is a probability measure defined on $\mathcal{A}_n$ by $\pi_n((A \cap C_n) \cup (B \cap C_n^c)) = \pi^*(A \cap C_n)$ for $A, B \in \mathcal{A}$ and $n \geq 1$. Consider the product $(S^{\mathbf{N}}, \prod_{n=1}^\infty \mathcal{A}_n, \prod_{n=1}^\infty \pi_n)$, then we have:

$$\big(\prod_{n=1}^\infty \pi_n\big)\big(\prod_{n=1}^\infty A_n\big) = \prod_{n=1}^\infty \pi_n(A_n) = \prod_{n=1}^\infty \pi(A_n) = \big(\pi^{\mathbf{N}}\big)\big(\prod_{n=1}^\infty A_n\big)$$

for all $A_n \in \mathcal{A}_n$ with $n \geq 1$. Hence we conclude $r\big(\prod_{n=1}^\infty \pi_n, \mathcal{A}^{\mathbf{N}}\big) = \pi^{\mathbf{N}}$, where $r(\,\cdot\,,\,\cdot\,)$ denotes the restriction of a measure to the σ-algebra. Thus for any $A \in \mathcal{A}^{\mathbf{N}}$ with $A \supset \prod_{n=1}^\infty C_n$ we have:

$$\pi^{\mathbf{N}}(A) = r\big(\prod_{n=1}^\infty \pi_n, \mathcal{A}^{\mathbf{N}}\big)(A) \geq \big(\prod_{n=1}^\infty \pi_n\big)\big(\prod_{n=1}^\infty C_n\big) = \prod_{n=1}^\infty \pi_n(C_n) = 1 \; .$$

This shows that $\big(\pi^{\mathbf{N}}\big)^*\big(\prod_{n=1}^\infty C_n\big) = 1$, and the proof of (2.131) is complete.

(2.132): First recall that for any random variable X we have (see [19] p.206):

(2.134) $$E|X| \leq 1 + \sum_{n=1}^\infty P\{\, |X| > n \,\} \; .$$

Next, under our hypothesis we claim that:

(2.135) $$\sum_{n=1}^\infty P\{\, |\eta_1| > nM \,\} < \infty$$

for some $M > 0$ large enough. For otherwise, we would have:

$$\sum_{n=1}^\infty P\{\, |\eta_1| > nm \,\} = \sum_{n=1}^\infty P\{\, |\eta_n|/n > m \,\} = \infty$$

for all $m \geq 1$. Hence by Borel-Cantelli's lemma (see Section 1.2) we get:

$$P\big(\limsup_{n\to\infty} \{|\eta_n|/n > m\}\big) = 1$$

for all $m \geq 1$. This readily implies:

$$P\big(\bigcap_{m=1}^\infty \{\limsup_{n\to\infty} |\eta_n|/n > m\}\big) = 1 \ .$$

It shows that $\limsup_{n\to\infty} |\eta_n|/n = +\infty$ P-a.s., and contradicts our hypothesis. Thus (2.135) is valid, and by (2.134) we get:

$$E|\eta_1/M| \leq 1 + \sum_{n=1}^\infty P\{\, |\eta_1/M| > n \,\} < \infty \ .$$

The proof of (2.132) is complete.

(2.133): Since $\pi^*(C) + \pi_*(C^c) = 1$ for all $C \subset S$, it is enough to show that:

$$\pi_*\{\, f < \alpha \ , \ \beta f < f^* \,\} = 0 \ .$$

For this, note that by definition of f^* we have:

$$\pi_*\{\, f < \alpha \, , \, \beta f < f^* \,\} \leq \pi_*\{\, f < \alpha \wedge (f^*)/\beta \,\} \leq \pi_*\{\, f \leq \alpha \wedge (f^*)/\beta < f^* \,\} = 0$$

since $\alpha \wedge (f^*)/\beta$ is π-measurable. The proof of (2.133) is complete.

We now pass to the main proof. Let us first consider the set:

$$C_n = \{\, \|f\|_T \geq n \,\} \cup \{\, 2\, \|f\|_T \geq \|f\|_T^* \,\}$$

for $n \geq 1$, where $\|f\|_T^*$ denotes the upper π-envelope of $\|f\|_T$ as a map from S to $\bar{\mathbf{R}}$. Then by (2.133) we have $\pi^*(C_n) = 1$ for all $n \geq 1$. Thus by (2.132) we have $\big(\pi^{\mathbf{N}}\big)^*\big(\prod_{n=1}^\infty C_n\big) = 1$. Now, let us consider the set:

$$C = \{\, (s_n)_{n\geq 1} \in S^{\mathbf{N}} \mid \tfrac{1}{n}\, \|f(s_n)\|_T \to 0 \,\} \ .$$

Since we have $\frac{1}{n} f(\xi_n) = \frac{1}{n}\sum_{j=1}^n f(\xi_j) - \frac{n-1}{n}\frac{1}{n-1}\sum_{j=1}^{n-1} f(\xi_j)$, then by our assumption (2.129) we may clearly conclude $\frac{1}{n}\, \|f(\xi_n)\|_T \to 0$ $(a.s.)$ as $n \to \infty$. Hence we easily find:

$$\big(\pi^{\mathbf{N}}\big)^*\big(C \cap \prod_{n=1}^\infty C_n\big) = 1 \ . \tag{2.136}$$

Finally, let us take $(s_n)_{n\geq 1} \in C \cap \prod_{n=1}^\infty C_n$. Then $\frac{1}{n}\, \|f(s_n)\|_T \to 0$ as $n \to \infty$, and thus there exists $n_0 \geq 1$ such that $\|f(s_n)\|_T < n$ for all $n \geq n_0$. Since $s_n \in C_n$, then $\|f(s_n)\|_T^* \leq 2\|f(s_n)\|_T$ for all $n \geq n_0$. This shows that $\frac{1}{n}\, \|f(s_n)\|_T^* \to 0$ as $n \to \infty$, for all $(s_n)_{n\geq 1} \in C \cap \prod_{n=1}^\infty C_n$. Thus by (2.136) we may conclude $\frac{1}{n}\, \|f(\xi_n)\|_T^* \to 0$ P-a.s., as $n \to \infty$. (It should be recalled that ξ_n is P-perfect (see (2.144) below), which implies $\|f\|_T^* \circ \xi_n = \|f(\xi_n)\|_T^*$ for $n \geq 1$.) Hence (2.130) follows by (2.132), and the proof is complete. □

2. In the remaining part of this section we show that the map f , which is subject to a uniform law of large numbers in the i.i.d. case, must be Gelfand π-integrable as a map from S into the Banach space $B(T)$. This is the second remarkable integrability consequence of the uniform law of large numbers. The result is found in [40] and requires the following definitions.

Let us for a moment assume that f is a map from S into a Banach space B with norm $\|\cdot\|$ and dual spaces B^* and B^{**} . (As above, we assume that $(S, \mathcal{A}, \pi)$ is a probability space.) Then f is said to be *weakly π-measurable*, if $\lambda \circ f$ is π-measurable for all $\lambda \in B^*$. The map f is said to be *weakly π-integrable*, if $\lambda \circ f$ is π-integrable for all $\lambda \in B^*$. In this case we define the *π-mean* of f to be a linear functional $\pi(f)$ on B^* defined by:

$$\pi(f)(\lambda) = \int_S \lambda(f)\, d\pi$$

for all $\lambda \in B^*$. It is easily verified that the linear operator $P : B^* \to L^1(\pi)$ defined by $P\lambda = \lambda(f)$ has a closed graph. Thus by the closed graph theorem (see [17] p.57) we find that P is bounded. It easily implies that $\pi(f) \in B^{**}$. The map f is said to be *Gelfand π-integrable*, if f is weakly π-integrable and $\pi(f)$ belongs to B . (As usual, we consider B as a closed subspace of B^{**}.) Thus, if f is Gelfand π-integrable, then there exists an element in B (denoted $\int_S f\ d\pi$ and called the *Gelfand π-integral* of f) such that:

$$\int_S \lambda(f)\, d\pi = \lambda\big(\int_S f\ d\pi\big)$$

for all $\lambda \in B^*$.

The Banach space of interest for us here is the space $B(T)$ with the sup-norm $\|\cdot\|_T$. (As before T is assumed to be an arbitrary set.) In this case it is easily verified that $f : S \to B(T)$ is Gelfand π-integrable if and only if the following three conditions are satisfied:

(2.137) $\quad t \mapsto M(t) = \int_S f(s,t)\ \pi(ds)$ exists as a map from T into $\mathbf{R}$ and belongs to $B(T)$

(2.138) $\quad s \mapsto \int_T f(s,t)\ \lambda(dt)$ is π-integrable as a map from S into $\mathbf{R}$, for all $\lambda \in ba(T)$

(2.139) $\quad \int_S \big(\int_T f(s,t)\ \lambda(dt) \big)\, \pi(ds) = \int_T \big(\int_S f(s,t)\ \pi(ds) \big)\, \lambda(dt)$ for all $\lambda \in ba(T)$.

It should be recalled that $ba(T)$ is the Banach space of all finitely additive real valued bounded functions defined on 2^T with the total variation on T as the norm. Moreover, it is well-known that $ba(T) = B(T)^*$ equals to the dual space of $B(T)$ (see [17]). The Gelfand π-integral of $f : S \to B(T)$ equals to M in (2.137). We are now ready to state the final result.

Theorem 2.26

Under the hypotheses of Theorem 2.25, let us assume that (2.129) is satisfied as $n \to \infty$ for $M \in B(T)$. Then $f : S \to B(T)$ is Gelfand π-integrable. (Thus (2.137)-(2.139) are also valid.)

Proof. Let $\lambda \in ba(T)$ be given and fixed. By (2.129) there is a P-null set $N \in \mathcal{F}$ such that:

$$\Big\| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j(\omega)) - M \Big\|_T \to 0 \tag{2.140}$$

as $n \to \infty$, for all $\omega \in \Omega \setminus N$. Now note that by Theorem 2.25 it is no restriction to assume that $f(\xi_j(\omega)) \in B(T)$ for all $\omega \in \Omega \setminus N$ and all $j \geq 1$. Thus, since λ is continuous on $B(T)$, from (2.140) we get:

$$\Big| \frac{1}{n} \sum_{j=1}^{n} \int_T f(\xi_j(\omega), t)\ \lambda(dt)\ - \int_T M(t)\, \lambda(dt) \Big| \to 0 \tag{2.141}$$

as $n \to \infty$, for all $\omega \in \Omega \setminus N$.

We establish that $\lambda \circ f$ is π-measurable by showing that $(\lambda \circ f)^* = (\lambda \circ f)_*$. For this, take any two π-measurable maps $g\,,h : S \to \bar{\mathbf{R}}$ satisfying $(\lambda \circ f)_* = g = h = (\lambda \circ f)^*$ on $\{\,(\lambda \circ f)_* = (\lambda \circ f)^*\,\}$, and $(\lambda \circ f)_* < g < h < (\lambda \circ f)^*$ on $\{\,(\lambda \circ f)_* < (\lambda \circ f)^*\,\}$. By definition of the π-envelopes we get $\pi_*\{\,g < \lambda \circ f\,\} \le \pi_*\{\,(\lambda \circ f)_* < g \le \lambda \circ f\,\} = 0$ and $\pi_*\{\,\lambda \circ f < h\,\} \le \pi_*\{\,\lambda \circ f \le h < (\lambda \circ f)^*\,\} = 0$. Hence we see that $\pi^*\{g \ge \lambda \circ f\} = \pi^*\{\lambda \circ f \ge h\} = 1$. Denote $L = \Omega \setminus N$, and put $L_0 = L \cap \{g \ge \lambda \circ f\}^{\mathbf{N}}$ and $L_1 = L \cap \{\lambda \circ f \ge h\}^{\mathbf{N}}$. Then by (2.131) we find $P^*(L_0) = P^*(L_1) = 1$. Now, first by definition of L_0 and (2.141) we get:

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} g(\xi_j(\omega)) \ge \int_T M(t)\ \lambda(dt)$$

for all $\omega \in L_0$. Second, by definition of L_1 and (2.141) we get:

$$\int_T M(t)\ \lambda(dt) \ge \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} h(\xi_j(\omega))$$

for all $\omega \in L_1$. Since the maps appearing in the last two inequalities are P-measurable, then these inequalities hold P-a.s. In other words, we obtain:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} g(\xi_j(\omega)) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} h(\xi_j(\omega)) = \int_T M(t)\ \lambda(dt)\quad P\text{-a.s.}$$

Hence from (2.132) we obtain $g\,,h \in L^1(\pi)$, and thus by Kolmogorov's Law 1.9 it follows:

$$\int_S g(s)\ \pi(ds) = \int_S h(s)\ \pi(ds) = \int_T M(t)\ \lambda(dt)\ . \tag{2.142}$$

Since $g \le h$, this implies $g = h$ π-a.s. Moreover, since $g < h$ on $\{\,(\lambda \circ f)_* < (\lambda \circ f)^*\,\}$, it follows $(\lambda \circ f)_* = (\lambda \circ f)^*$. Thus $\lambda \circ f$ is π-measurable.

Finally, since $g = h = \lambda \circ f$ π-a.s., from (2.142) we get:

$$\int_S (\lambda \circ f)\ d\pi = \int_T \Big(\int_S f(s,t)\ \pi(ds) \Big)\ \lambda(dt) = \lambda\Big(\int_S f\ d\pi \Big)\ .$$

This shows that $f : S \to B(T)$ is Gelfand π-integrable. The proof is complete. $\square$

2.5.3 Necessity of measurability. The purpose of this section is to show that a measurability of the underlying map f is a necessary condition for the uniform law of large numbers to hold in the i.i.d. case. The result appears in [1] and is due to Talagrand [84].

Let $(S, \mathcal{A}, \pi)$ be a probability space, and let $(\Omega, \mathcal{F}, P)$ be the countable product $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}}, \pi^{\mathbf{N}})$. Let $\xi_j\big((s_i)_{i\ge1}\big) = s_j$ be the j-th projection of $S^{\mathbf{N}}$ onto S for $(s_i)_{i\ge1} \in S^{\mathbf{N}}$ and $j \ge 1$. Then $\{\xi_j\}_{j\ge1}$ is a sequence of independent and identically distributed random variables defined on $(\Omega, \mathcal{F}, P)$ with values in $(S, \mathcal{A})$ and common distribution law π . Let T be a set, and let $f : S \times T \to \mathbf{R}$ be a given map. We denote $S_n(f) = \sum_{j=1}^{n} f(\xi_j)$ for all $n \ge 1$. In the sequel we adopt the notation from Section 2.2. The main result of this section may now be stated as follows.

Theorem 2.27

Let $\{\xi_j\}_{j\geq 1}$ *be a sequence of independent and identically distributed random variables defined on* $(\Omega, \mathcal{F}, P)$ *as above. Let us assume that:*

$$\big\| \tfrac{1}{n} S_n(f) - M \big\|_T \to 0 \quad (P^*) \tag{2.143}$$

as $n \to \infty$ *for* $M \in B(T)$ *. Then* $s \mapsto f(s,t)$ *is* π*-measurable for all* $t \in T$ *, as a map from* S *into* $\mathbf{R}$ *.*

Proof. The proof is based upon the following two facts:

(2.144) Each ξ_n is π-perfect, and thus $f(\xi_n, t)^* = f(\,\cdot\,,t)^* \circ \xi_n$ for all $n \geq 1$ and all $t \in T$

(2.145) We have $\big(\sum_{j=1}^n f(\xi_j, t)\big)^* = \sum_{j=1}^n f(\xi_j, t)^*$ for all $n \geq 1$ and all $t \in T$.

The arguments leading to (2.144)-(2.145) may be presented as follows.

(2.144): According to the classic characterization of perfectness (see [62]), it is enough to show that $\pi^*(A) = 1$ implies $(\pi^{\mathbf{N}})^*(S_1 \times \ldots \times S_{n-1} \times A \times S_{n+1} \times \ldots\,) = 1$ with $S_j = S$ for $j \geq 1$ and $A \in \mathcal{A}$. This fact follows clearly from (2.131), and the proof of (2.144) is complete.

(2.145): We will assume that $n = 2$. It is easily verified that the same arguments apply in the general case as well. Let $t \in T$ be given and fixed. Note that $g_n \uparrow g$ implies $g_n^* \uparrow g^*$, and therefore it is no restriction to assume that $f(\,\cdot\,,t)$ is bounded from above by a constant. Thus $f(\,\cdot\,,t)^* < \infty$, and $f(\,\cdot\,,t)^* - \varepsilon$ is a well-defined function for all $\varepsilon > 0$. The proof is essentially based on the following fact:

(2.146) If for some $C \in \mathcal{A} \otimes \mathcal{A}$ we have $(\pi \otimes \pi)(C) > 0$, then $\pi\{\, s \in S \mid \pi(C_s) > 0 \,\} > 0$

where $C_s = \{\, u \in S \mid (s,u) \in C \,\}$ denotes the section of C at $s \in S$. It follows from the well-known formula on the product measures $(\pi \otimes \pi)(C) = \int_S \pi(C_s)\; \pi(ds)$ (see [19]).

Now we pass to the proof itself. First, note that by definition of the upper envelope we get:

$$\big(f(\xi_1, t) + f(\xi_2, t)\big)^* \leq f(\xi_1, t)^* + f(\xi_2, t)^* \ .$$

Thus by (2.144) it is enough to show that:

$$\big(f(\xi_1, t) + f(\xi_2, t)\big)^*(s_1, s_2) \geq f(s_1, t)^* + f(s_2, t)^* \tag{2.147}$$

for $(\pi \otimes \pi)$-a.s. $(s_1, s_2) \in S \otimes S$. (To be more precise, it should be noted that here we use the fact that $(\xi_1, \xi_2) : S^{\mathbf{N}} \to S \otimes S$ is $(\pi \otimes \pi)$-perfect, which is easily seen as for (1.144) via an easy generalization of (2.131) where the product space may consist of different factors. This is the reason for which the upper $\pi^{\mathbf{N}}$-envelope on the left hand side in (2.147) depends only upon the first two coordinates.) For (2.147) put $M(s_1, s_2) = \big(f(\xi_1, t) + f(\xi_2, t)\big)^*(s_1, s_2)$ for $(s_1, s_2) \in S \otimes S$, and suppose that there exist $n, m \geq 1$ such that:

$$(\pi \otimes \pi)\big\{\, (s_1, s_2) \in S \otimes S \mid M(s_1, s_2) < (f(s_1, t)^* - 2^{-n}) + (f(s_2, t)^* - 2^{-m}) \,\big\} > 0 \ . \tag{2.148}$$

It is no restriction to assume that $M \geq f(\xi_1, t) + f(\xi_2, t)$ on $S \otimes S$. Then by (2.146) we have:

$$\pi\big\{\, s_1 \in S \mid \pi\{\, s_2 \in S \mid M(s_1, s_2) < (f(s_1, t)^* - 2^{-n}) + (f(s_2, t)^* - 2^{-m}) \,\} > 0 \,\big\} > 0 \ . \tag{2.149}$$

Denote by A the set of all $s_1 \in S$ appearing in (2.149). Then there exists $\hat{s}_1 \in A$ such that:

$$f(\hat{s}_1,t)^* - 2^{-n} < f(\hat{s}_1,t) \ . \tag{2.150}$$

For otherwise, we would have $A \subset \{ \, s_1 \in S \mid f(s_1,t) \le f(s_1,t)^* - 2^{-n} < f(s_1,t)^* \, \}$, and thus $\pi(A) = 0$ by definition of the upper envelope. It contradicts (2.149), and thus (2.150) holds. Finally, for $\hat{s}_1 \in S$ from (2.150) we get:

$$\begin{aligned} 0 &< \pi_*\{ \, s_2 \in S \mid f(\hat{s}_1,t) + f(s_2,t) \le M(\hat{s}_1,s_2) < (f(\hat{s}_1,t)^* - 2^{-n}) + (f(s_2,t)^* - 2^{-m}) \, \} \\ &\le \pi_*\{ \, s_2 \in S \mid f(s_2,t) \le f(s_2,t)^* - 2^{-m} < f(s_2,t)^* \, \} \ . \end{aligned}$$

This is not possible by definition of the upper envelope. Thus (2.148) fails for all $n, m \ge 1$, and from this fact we easily obtain (2.147). The proof of (2.145) is complete.

We turn to the main proof. First we recall the following well-known technical facts:

$$P^*\{ g > \varepsilon \} = P\{ g^* > \varepsilon \} \tag{2.151}$$

$$|\, g^* - h^* \,| \le |\, g - h \,|^* \tag{2.152}$$

$$|\, g_* - h_* \,| \le |\, g - h \,|^* \tag{2.153}$$

where the objects are self-explained, and provided that the left-hand sides in (2.152) and (2.153) are well-defined. Now by (2.143), (2.144), (2.145), (2.151), (2.152) and (2.153) we find:

$$\frac{1}{n} \sum_{j=1}^{n} f(\,\cdot\,,t)^* \circ \xi_j \to M(t) \quad \text{in } P\text{-probability} \tag{2.154}$$

$$\frac{1}{n} \sum_{j=1}^{n} f(\,\cdot\,,t)_* \circ \xi_j \to M(t) \quad \text{in } P\text{-probability} \tag{2.155}$$

as $n \to \infty$. Hence we get $\pi\{ \, |f(\,\cdot\,,t)^*| = +\infty \, \} = \pi\{ \, |f(\,\cdot\,,t)_*| = +\infty \, \} = 0$, and thus:

$$\frac{1}{n} \sum_{j=1}^{n} \big(f(\,\cdot\,,t)^* - f(\,\cdot\,,t)_* \big) \circ \xi_j \longrightarrow 0 \quad \text{in } P\text{-probability} \tag{2.156}$$

as $n \to \infty$. The sequence $\big\{ \, (f(\,\cdot\,,t)^* - f(\,\cdot\,,t)_*) \circ \xi_j \, \big\}_{j \ge 1}$ is an i.i.d. sequence of non-negative random variables. Thus from (2.156) by Kolmogorov's Law 1.9 we get:

$$\frac{1}{n} \sum_{j=1}^{n} \Big(\big(f(\,\cdot\,,t)^* - f(\,\cdot\,,t)_* \big) \circ \xi_j \Big) \wedge N \longrightarrow 0 \quad P\text{-a.s.}$$

as $n \to \infty$, for all $N \ge 1$. It moreover implies that:

$$\int_{S^{\mathbf{N}}} \Big(\big(f(\,\cdot\,,t)^* - f(\,\cdot\,,t)_* \big) \circ \xi_1 \Big) \wedge N \; d\big(\pi^{\mathbf{N}}\big) = 0$$

for all $N \ge 1$. Letting $N \to \infty$ we obtain:

$$\int_{S^{\mathbf{N}}} \Big(\big(f(\,\cdot\,,t)^* - f(\,\cdot\,,t)_* \big) \circ \xi_1 \Big) \; d\big(\pi^{\mathbf{N}}\big) = \int_S \big(f(s,t)^* - f(s,t)_* \big) \; \pi(ds) = 0 \ .$$

Thus $f(\,\cdot\,,t)^* = f(\,\cdot\,,t)_*$ π-a.s., and the proof is complete. $\square$

2.5.4 Gerstenhaber's counter-example. In this section we present a counter-example of M. Gerstenhaber (as recorded in [36] p.32) which shows that Theorem 2.25 and Theorem 2.27 (from the previous two sections) may fail in the general stationary ergodic case. The method is well-known and has wide applications in the construction of various examples in ergodic theory. We first explain the meaning of the result in more detail.

First, let us mention that the index set T will consist of a single point. Second, note that from (2.132) we get $\xi_1 \in L^1(P)$ whenever $\{\xi_j\}_{j\geq 1}$ is a sequence of independent and identically distributed random variables defined on $(\Omega, \mathcal{F}, P)$ satisfying:

$$\frac{1}{n}\sum_{j=1}^{n} \xi_j \to C \quad P\text{-a.s.} \tag{2.157}$$

as $n \to \infty$. This follows readily from the fact that $\frac{1}{n}|\xi_n| = |\frac{1}{n}\sum_{j=1}^{n}\xi_j - \frac{n-1}{n}\frac{1}{n-1}\sum_{j=1}^{n-1}\xi_j| \to 0$ P-a.s. for $n \to \infty$. Third, this fact extends to the stationary ergodic case if ξ_j's are assumed to be non-negative. Indeed, if $\{\xi_j\}_{j\geq 1}$ is stationary and ergodic satisfying (2.157) with $\xi_j \geq 0$ P-a.s. for $j \geq 1$, then by (2.10)-(2.12) we get:

$$\frac{1}{n}\sum_{j=1}^{n}(\xi_j \wedge N) \to E(\xi_1 \wedge N) \quad P\text{-a.s.}$$

as $n \to \infty$, for all $N \geq 1$. Since $\xi_j \wedge N \leq \xi_j$ for all $j \geq 1$, we get from (2.157) that $E(\xi_1 \wedge N) \leq C$ for all $N \geq 1$. Letting $N \to \infty$ we obtain $\xi_1 \in L^1(P)$. Fourth, this result obviously extends to the case where we know that either $\xi_1^+ \in L^1(P)$ or $\xi_1^- \in L^1(P)$. Finally, the next example shows that the result may fail in general. In other words, we show that (2.157) can be valid (with $C = 0$) even though $\xi_1 \notin L^1(P)$ (actually $E(\xi_1^+) = E(\xi_1^-) = +\infty$) . Furthermore, a slight modification of the example will show that Theorem 2.27 may fail in the stationary ergodic case. More precisely, we show that:

$$\frac{1}{n}\sum_{j=1}^{n} f(\xi_j) \to C \quad (a.s.)^* \tag{2.158}$$

as $n \to \infty$, even though f is not π-measurable. (As usual π denotes the distribution law of ξ_1.)

Example 2.28 (Gerstenhaber)

Let Φ_0 be an invertible ergodic (measure-preserving) transformation of $X_0 = [0,1[$, equipped with the Borel σ-algebra $\mathcal{B}(X_0)$ and Lebesgue measure λ . (For instance, the transformation Φ_0 could be an irrational rotation in $[0,1[$ defined by $\Phi_0(x) = x+\alpha \ (mod\ 1)$ for some $\alpha \in [0,1[\setminus\mathbf{Q}$ with $x \in X_0$. For more details see Section 1.1.) Let $\{a_j\}_{j\geq 0}$ be a sequence of real numbers satisfying $1 = a_0 \geq a_1 \geq \ldots \geq 0$, let $X_n = [0,a_n[\times\{n\}$ for all $n \geq 0$, and let $X = \cup_{n=0}^{\infty} X_n$. The σ-algebra $\mathcal{B}$ on X will consist of those $B \subset X$ which satisfy $p_1(B\cap X_n) \in \mathcal{B}(X_0)$ for all $n \geq 0$. (Here $p_1 : \mathbf{R}^2 \to \mathbf{R}$ denotes the projection onto the first coordinate.) The measure μ on $\mathcal{B}$ is defined by $\mu(B) = \sum_{n=0}^{\infty}\lambda(p_1(B\cap X_n))$ for all $B \in \mathcal{B}$. The transformation Φ of X is defined by $\Phi(x,y) = (x,y+1)$ if $x < a_{y+1}$, and $\Phi(x,y) = (\Phi_0(x),0)$ if $x \geq a_{y+1}$, whenever $(x,y) \in X$. Then it is easily verified that Φ is an invertible ergodic (measure-preserving) transformation in the measure space $(X, \mathcal{B}, \mu)$.

Now we choose the sequence $\{a_j\}_{j\geq 0}$ in a particular way described by the conditions:

$$a_1 = a_2 \ , \ a_3 = a_4 \ , \ a_5 = a_6 \ , \ \ldots \tag{2.159}$$

$$\textstyle\sum_{n=1}^{\infty} a_{2n} < \infty \tag{2.160}$$

$$\textstyle\sum_{n=1}^{\infty} \sqrt{n}\, a_{2n} = \infty \ . \tag{2.161}$$

(For instance, the choice $a_{2n} = n^{-3/2}$ for $n \geq 1$ suffices.) Note that from (2.159)+(2.160) we have $\mu(X) = \sum_{n=0}^{\infty} a_n < \infty$.

Next we define a map $F : X \to \mathbf{R}$. We put $F(z) = 0$ for $z \in X_0$, $F(z) = -\sqrt{n}$ for $z \in X_{2n-1}$, and $F(z) = \sqrt{n}$ for $z \in X_{2n}$, whenever $n \geq 1$. Then by construction it is easily verified that we have:

$$\Big| \sum_{j=0}^{n-1} F(\Phi^j) \Big| \leq \sqrt{n/2} \tag{2.162}$$

for all $n \geq 1$. Putting $\xi_j = F(\Phi^{j-1})$ for $j \geq 1$, we find by (2.9) that $\{\xi_j\}_{j\geq 1}$ is stationary and ergodic. Moreover, by (2.159)+(2.161) we clearly have $E(\xi_1^{\pm}) = \int_X F^{\pm}\, d\mu = +\infty$. Finally, from (2.162) we obtain $\frac{1}{n}\sum_{j=1}^{n} \xi_j \to 0$ everywhere on X , as $n \to \infty$. Thus (2.157) is satisfied, and our first claim stated above is complete.

We turn to the second claim by modifying definition of the map F . Namely, if we take any set $C \subset [0, a_1[$ which is not λ-measurable, and define $f(z) = F(z) \cdot 1_{C\times\{n\}}(z)$ for all $z \in X_n$ and all $n \geq 0$, then as for (2.162) we get:

$$\Big| \sum_{j=1}^{n} f(\xi_j) \Big| \leq \sqrt{n/2} \tag{2.163}$$

for all $n \geq 1$, where $\xi_j = \Phi^{j-1}$ for $j \geq 1$. By (2.9) we know that $\{\xi_j\}_{j\geq 1}$ is stationary and ergodic. Moreover, from (2.163) we get (2.158) with the null-set being empty. Finally, it is obvious that $f : X \to \mathbf{R}$ is not μ-measurable (for example $f^{-1}(\{-1\}) = C \times \{1\} \notin \mathcal{B}_\mu$) . These facts complete the second claim stated above.

2.5.5 A uniform ergodic theorem over the orbit – Weyl's theorem. In the next example we show that if the unit ball of a Banach space (which appears in the Yosida-Kakutani uniform ergodic theorem in Section 1.3) has been replaced by a smaller family of vectors, the uniform ergodic theorem is much easier established, but still of considerable interest.

Example 2.29 (A uniform ergodic theorem over the orbit – Weyl's theorem)

Consider the d-dimensional torus $X = [0,1[^d$ equipped with the d-dimensional Lebesgue measure $\mu = \lambda_d$ for some $d \geq 1$. Then X is a compact group with respect to the coordinatewise addition mod 1. Take a point $\alpha = (\alpha_1, \ldots, \alpha_d)$ in X such that $\alpha_1, \ldots, \alpha_d$ and 1 are rationally independent, which means if $\sum_{i=1}^{d} k_i \alpha_i$ is an integer for some integers $k_1, \ldots, k_d$, then $k_1 = \ldots = k_d = 0$. Under this condition the translation $T(x) = x + \alpha$ is an ergodic (measure-preserving) transformation of X , and the orbit $\mathcal{O}(x) = \{\, T^j(x) \mid j \geq 0 \,\}$ is dense in X for every $x \in X$ (see [52] p.12). Let $f : X \to \mathbf{R}$ be a continuous function, and let $\mathcal{O}(f) = \{\, T^j(f) \mid j \geq 0 \,\}$ be the orbit of f under T . Then the uniform ergodic theorem (1.1) over the orbit $\mathcal{F} = \mathcal{O}(f)$ as $n \to \infty$ is stated as follows:

$$\sup_{g \in \mathcal{O}(f)} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(g) - \int f \, d\mu \Big| \to 0 \tag{2.164}$$

To verify its validity, we shall first note that (2.164) is equivalently written as follows:

$$\sup_{i \geq 0} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^{i+j}(f) - \int f \, d\mu \Big| \to 0 \tag{2.165}$$

for $n \to \infty$. Then, by the continuity of f and the denseness of $\mathcal{O}(x)$ for every $x \in X$, we see that (2.165) is equivalent to the following fact:

$$\sup_{x \in X} \Big| \frac{1}{n} \sum_{j=0}^{n-1} T^j(f)(x) - \int f \, d\mu \Big| \to 0 \tag{2.166}$$

for $n \to \infty$. However, this is precisely the content of *Weyl's theorem* [91] *on uniform distribution mod 1* (see [52] p.13), and thus the uniform ergodic theorem (2.164) is established.

The Weyl's result extends to general compact metric spaces: If T is an ergodic isometry of a compact metric space X , then T is *uniquely ergodic* (there is exactly one T-invariant Borel probability measure on X) if and only if, for every continuous real valued map f on X , the ergodic averages of f converge uniformly to a constant (see [70] p.137-139 and [52] p.12-13). From this general fact one can derive a number of corollaries (such as a uniform ergodic theorem for any family $\{ f \circ S \mid S$ is an ergodic isometry of $X \}$ where f is a continuous real valued map on X). This kind of example also raises the question about general connections between uniform ergodic theorems and uniqueness of certain invariant linear functionals on the linear span of the relevant family of functions $\mathcal{F}$. These aspects are to a certain extent well described in the literature, and the interested reader is referred to [70] (p.135-150) where more details can be found (a connection with *almost periodic sequences* for instance).

2.5.6 Metric entropy and majorizing measure type conditions. In this section we present several (unsuccessful) trials to apply metric entropy and majorizing measure type conditions (which emerged in the study of regularity (boundedness and continuity) of random processes) for obtaining the uniform law of large numbers. In this process we realize that the main obstacle to a more successful application of these conditions (which are known to be sharp in many ways) towards the uniform law of large numbers relies upon the fact that they imply a continuity of the underlying process (actually sequence of processes), while the uniform law of large numbers requires (and is actually equivalent to) an asymptotic continuity. (A more detailed explanation of this phenomenon will not be given here.) We start by explaining the main observation which motivated our attempts mentioned above.

1. Throughout we assume that $\xi = \{ \xi_j \mid j \geq 1 \}$ is a sequence of independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$ (which will mostly be just the real line with Borel σ-algebra) and the common distribution law π . We suppose that a set T is given, as well as a map $f : S \times T \to \mathbf{R}$ such that $s \mapsto f(s,t)$ is π-measurable for every $t \in T$. We put $M(t) = \int_S f(s,t) \, \pi(ds)$ to

denote the π-mean function of f for $t \in T$. In this context our main concern is the uniform law of large numbers which states:

$$\sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j, t) - M(t) \right| \to 0 \ \ (a.s.)^* \ \& \ (L^1)^* \tag{2.167}$$

as $n \to \infty$. By Corollary 2.8 we know that (2.167) is equivalent (provided that $\|f\|_T \in L^{<1>}(\pi)$ and $\xi : \Omega \to S^{\mathbf{N}}$ is P-perfect) to the fact that f is eventually totally bounded in π-mean, which according to (2.29) means:

(2.168) $\forall \varepsilon > 0$, $\exists \gamma_\varepsilon \in \Gamma(T)$ such that:

$$\inf_{n \geq 1} \frac{1}{n} \int^* \sup_{t', t'' \in A} \left| S_n\big(f(t') - f(t'')\big) \right| dP < \varepsilon$$

for all $A \in \gamma_\varepsilon$.

(We recall that $\Gamma(T)$ denotes the family of all finite coverings of the set T .) Thus, the main task in verifying (2.168) (and thus (2.167) as well) is to estimate the expression:

$$\int^* \sup_{t', t'' \in A} \left| \tfrac{1}{n} S_n\big(f(t') - f(t'')\big) \right| dP \tag{2.169}$$

for $A \subset T$ and $n \geq 1$. Putting $X_t^n = \frac{1}{n} S_n\big(f(t)\big)$ for $t \in T$ and $n \geq 1$ (and passing over measurability problems) we come to the problem of estimating the expression:

$$E \sup_{s, t \in A} |X_s^n - X_t^n| \tag{2.170}$$

for $A \subset T$ and $n \geq 1$.

2. On the other hand in the study of regularity of random processes (under metric entropy conditions) we have the following basic result (see [54] p.300):

Theorem A. *Let $X = (X_t)_{t \in T}$ be a random process in $L^\psi(P)$ satisfying the condition:*

$$\|X_s - X_t\|_\psi \leq d(s, t) \tag{2.171}$$

for all $s, t \in T$. Then we have:

$$E \sup_{s, t \in T} |X_s - X_t| \leq 8 \int_0^D \psi^{-1}\big(N(T, d; \varepsilon)\big) \, d\varepsilon \ . \tag{2.172}$$

Here ψ denotes a *Young function* ($\psi : \mathbf{R}_+ \to \mathbf{R}_+$ is convex, increasing, and satisfies $\psi(0) = 0$ and $\psi(+\infty) = +\infty$), and $L^\psi(P)$ denotes the *Orlicz space* associated with ψ (it consists of random variables Z on $(\Omega, \mathcal{F}, P)$ satisfying $\|Z\|_\psi < \infty$, where $\|\cdot\|_\psi$ is the *Orlicz norm* associated with ψ defined by $\|Z\|_\psi = \inf\{\, c > 0 \mid E\psi(|Z|/c) \leq 1 \,\}$). The letter d denotes a pseudo-metric on T , and D is the diameter of (T, d) . The symbol $N(T, d; \varepsilon)$ denotes the *entropy number*, which is the smallest number of open balls of radius $\varepsilon > 0$ in the pseudo-metric d needed to cover T (see Definition in the beginning of this chapter). Note that without loss of generality the set T is assumed to be finite (countable).

It is moreover well-known that condition (2.171) in Theorem A (with $\psi(x) = x^p$ for $1 < p < \infty$) may be equivalently replaced by the seemingly weaker condition:

$$\|X_s - X_t\|_{p,\infty} \leq d(s,t) \tag{2.173}$$

for all $s,t \in T$. Here $\|\cdot\|_{p,\infty}$ denotes the weak L^p-norm defined as follows $\|Z\|_{p,\infty} = \sup_{t>0}(t^p P\{\,|Z| > t\,\})^{1/p}$ whenever Z is a random variable on $(\Omega, \mathcal{F}, P)$.

Finally, in the study of regularity of random processes (under majorizing measure conditions) we have the following basic result (see [54] p.316):

Theorem B. *Let $X = (X_t)_{t\in T}$ be a random process in $L^\psi(P)$ satisfying condition (2.171). Assume moreover that ψ satisfies $\psi^{-1}(x\,y) \leq C\,(\psi^{-1}(x) + \psi^{-1}(y))$ for all $x,y > 0$ with some $C > 0$, and $\int_0^1 \psi^{-1}(x^{-1})\;dx < \infty$. (For instance $\psi(x) = \exp(x^p) - 1$ with $1 \leq p < \infty$ is of that type.) Then we have:*

$$E \sup_{s,t\in T} |X_s - X_t| \leq K_\psi \cdot \sup_{t\in T} \int_0^D \psi^{-1}\Big(\frac{1}{m(B_d(t,\varepsilon))}\Big)\;d\varepsilon \tag{2.174}$$

for any probability measure m on (T,d) . (Here K_ψ is a numerical constant, and $B_d(t,\varepsilon)$ denotes the ball with center at $t \in T$ and radius $\varepsilon > 0$ in the pseudo-metric d .)

3. Now, having in mind estimates (2.172) and (2.174), one may naturally ask if we could use these to bound the expression in (2.170) in such a way that the integrals involved are finite. This would clearly lead to (2.168) (by splitting T into finitely many subsets of small diameters which would make the integrals involved desirably small). Actually, it should be noted that in the case where we apply (2.174) to bound (2.170) towards (2.168), we need the following stronger condition:

$$\lim_{\delta\downarrow 0} \sup_{t\in T} \int_0^\delta \psi^{-1}\Big(\frac{1}{m(B_d(t,\varepsilon))}\Big)\;d\varepsilon = 0\ .$$

This requirement turns out to distinguish the continuity from the boundedness of (a version of) the process $X = (X_t)_{t\in T}$ under majorizing measure conditions (see [54]).

4. Thus, the procedure of applying (2.172) and (2.174) to (2.170) may now be described more precisely as follows. First, we should have fulfilled a separability assumption on the process $(X_t^n)_{t\in T}$ for every $n \geq 1$, which would reduce the supremum in (2.170) over a countable set. Second, we consider the incremental condition:

$$\|X_s^n - X_t^n\|_\psi \leq d_n(s,t) \tag{2.175}$$

for all $s,t \in T$ and all $n \geq 1$, where d_n is a pseudo-metric on T for $n \geq 1$. In the case where $\psi(x) = x^p$ for $1 < p < \infty$, we can relax (2.175) to the condition:

$$\|X_s^n - X_t^n\|_{p,\infty} \leq d_n(s,t) \tag{2.176}$$

for all $s,t \in T$ and all $n \geq 1$. Finally, we look at the integrals involved and require:

$$\inf_{n\geq 1} \int_0^{D_n} \psi^{-1}\big(N(T,d_n;\varepsilon)\big)\;d\varepsilon < \infty \tag{2.177}$$

where D_n is the diameter of (T, d_n) for $n \geq 1$. Similarly, we require:

$$\lim_{\delta \downarrow 0} \sup_{t \in T} \int_0^\delta \psi^{-1}\Big(\frac{1}{m(B_{d_n}(t,\varepsilon))} \Big)\, d\varepsilon = 0 \tag{2.178}$$

for some $n \geq 1$ and some probability measure m on (T, d_n) .

From the arguments and results presented above we see that either (2.177) or (2.178) is sufficient for (2.168), and thus for (2.167) as well. However, although the procedure just described seems reliable at first sight, as a matter of fact, it does not work. Our main aim in this section is to illustrate this fact by means of various examples in which this procedure has been unsuccessfully repeated. We postpone to clarify this interesting phenomenon in more detail elsewhere (through the concept of an asymptotic continuity of random processes which we realize to be the key point).

5. The examples below make use of the notation which appears throughout the whole section. Common to them is the i.i.d. sequence $\{ \xi_j \mid j \geq 1 \}$ defined on $(\Omega, \mathcal{F}, P)$ with values in $\mathbf{R}$ and distribution law π . The set T equals to $\mathbf{R}$, and the map f is given by $f(s,t) = 1_{<-\infty,t]}(s)$ for all $(s,t) \in \mathbf{R}^2$. We moreover have:

$$X_t^n = \frac{1}{n} \sum_{j=1}^n 1_{\{\xi_j \leq t\}}$$

for $t \in T$ and $n \geq 1$. Hence we find:

$$|X_s^n - X_t^n| \sim \tfrac{1}{n} B(n, |F(s) - F(t)|)$$

for all $s,t \in T$ and all $n \geq 1$. Here $F(t) = P\{\xi_1 \leq t\}$ is the distribution function of ξ_1 , and $B(n,p)$ denotes the binomial distribution with parameters $n \geq 1$ and $p \in [0,1]$. Below we denote $G(s,t) = |F(s) - F(t)|$ for $s,t \in T$. We first direct our attention to the condition (2.177).

Example 2.30

The example consists of five cases as follows.

1° Let $\psi(x) = x$ for $x \geq 0$. Then we have $\|X_s^n - X_t^n\|_\psi = |F(s) - F(t)| := d_n(s,t)$, and thus $N(T, d_n; \varepsilon) \sim N([0,1], |\cdot|; \varepsilon) \sim \frac{1}{\varepsilon}$. Since $\psi^{-1}(y) = y$, hence we get:

$$\int_0^{D_n} \psi^{-1}\big(N(T, d_n; \varepsilon)\big)\, d\varepsilon = \int_0^1 \frac{1}{\varepsilon}\, d\varepsilon = +\infty$$

for all $n \geq 1$. Thus (2.177) fails in this case.

2° Let $\psi(x) = x^2$ for $x \geq 0$. Then we have $\|X_s^n - X_t^n\|_\psi = \big(\frac{1}{n} G(s,t) + (1 - \frac{1}{n}) G(s,t)^2\big)^{1/2} := d_n(s,t) \geq (\frac{1}{n}\, G(s,t))^{1/2} := \delta_n(s,t)$. Thus $N(T, d_n; \varepsilon) \geq N(T, \delta_n; \varepsilon) \sim N([0,1], (\frac{1}{n}|\cdot|)^{1/2}; \varepsilon) \sim \frac{1}{n\varepsilon^2}$. Since $\psi^{-1}(y) = \sqrt{y}$, hence we get:

$$\int_0^{D_n} \psi^{-1}\big(N(T, d_n; \varepsilon)\big)\, d\varepsilon \geq \int_0^{\frac{1}{\sqrt{n}}} \psi^{-1}\big(N(T, \delta_n; \varepsilon)\big)\, d\varepsilon$$

$$= \int_0^{\frac{1}{\sqrt{n}}} \psi^{-1}\Big(\frac{1}{n\varepsilon^2}\Big)\, d\varepsilon = \frac{1}{\sqrt{n}} \int_0^{\frac{1}{\sqrt{n}}} \frac{1}{\varepsilon}\, d\varepsilon = +\infty$$

for all $n \geq 1$. Thus (2.177) fails in this case.

3° Let $\psi(x) = \exp(x^2) - 1$ for $x \ge 0$, and let $n = 1$. Then we have $\|X_s^1 - X_t^1\|_\psi = 1/\sqrt{\log(1+1/G(s,t))} := d_1(s,t)$. Thus we get $N(T, d_1; \varepsilon) \sim N([0,1], |\cdot|; 1/(\exp(1/\varepsilon^2) - 1)) \sim \exp(1/\varepsilon^2) - 1$. Since $\psi^{-1}(y) = \sqrt{\log(1+y)}$, hence we get:

$$\int_0^{D_1} \psi^{-1}\big(N(T, d_1; \varepsilon)\big)\, d\varepsilon = \int_0^{1/\sqrt{\log 2}} \frac{1}{\varepsilon}\, d\varepsilon = +\infty .$$

Thus (2.177) fails in this case with the infimum being attained for $n = 1$.

4° Let $\psi(x) = \exp(x) - 1$ for $x \ge 0$. Then in order to compute $\|X_s^n - X_t^n\|_\psi$, we look at $E\exp(|X|/C)$, where $X \sim B(n,p)$ with $p = G(s,t)$. We have:

$$\begin{aligned} E\exp(|X|/C) &= \sum_{k=0}^{n} \exp(k/C) \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} \big(p\cdot\exp(1/C)\big)^k (1-p)^{n-k} = \big(p\cdot\exp(1/C) + (1-p)\big)^n = 2 \end{aligned}$$

if and only if $C = 1/\log(1+(2^n-1)/G(s,t))$. Thus $\|X_s^n - X_t^n\|_\psi = 1/(n\log(1+(2^n-1)/G(s,t))) := d_n(s,t)$, and therefore $N(T, d_n; \varepsilon) \sim N([0,1], |\cdot|; (2^n-1)/(\exp(1/n\varepsilon) - 1)) \sim (\exp(1/n\varepsilon) - 1)/(2^n - 1)$. Since $\psi^{-1}(y) = \log(1+y)$, hence we get:

$$\begin{aligned} \int_0^{D_n} \psi^{-1}\big(N(T, d_n; \varepsilon)\big)\, d\varepsilon &= \int_0^{D_n} \log\Big(1 + \frac{1}{2^n - 1}\big(\exp(1/n\varepsilon) - 1\big)\Big)\, d\varepsilon \\ &\ge \int_0^{D_n} \log\Big(\frac{1}{2^n - 1}\big(\exp(1/n\varepsilon) - 1\big)\Big)\, d\varepsilon = \int_0^{D_n} \Big(\frac{1}{n\varepsilon} - \log\Big(\frac{1}{2^n - 1}\Big)\Big)\, d\varepsilon \\ &= \frac{1}{n} \int_0^{D_n} \frac{1}{\varepsilon}\, d\varepsilon - D_n \log\Big(\frac{1}{2^n - 1}\Big) = +\infty \end{aligned}$$

for all $n \ge 1$. Thus (2.177) fails in this case.

5° Let $\psi(x) = x^p$ for $x \ge 0$ with $1 < p < \infty$, and let $n = 1$. Then $\|X_s^1 - X_t^1\|_{p,\infty} = G(s,t)^{1/p} := d_1(s,t)$. Thus $N(T, d_1; \varepsilon) \sim N([0,1], |\cdot|; \varepsilon^p) \sim \frac{1}{\varepsilon^p}$. Since $\psi^{-1}(y) = y^{1/p}$, we get:

$$\int_0^{D_1} \psi^{-1}\big(N(T, d_1; \varepsilon)\big)\, d\varepsilon = \int_0^{1} \frac{1}{\varepsilon}\, d\varepsilon = +\infty .$$

Thus (2.177) fails in this case with the infimum being attained for $n = 1$.

Now we turn to the condition (2.178). Again, we denote $G(s,t) = |F(s) - F(t)|$ for $s, t \in T$.

Example 2.31

The example consists of two cases as follows.

1° Let $\psi(x) = \exp(x^2) - 1$ for $x \ge 0$, and let $n = 1$. Then we have $\|X_s^1 - X_t^1\|_\psi = 1/\sqrt{\log(1 + 1/G(s,t))} := d_1(s,t)$. Let $F \sim U(0,1)$ be the function of uniform distribution on $[0,1]$. Thus $F(t) = t$ for $t \in [0,1]$, $F(t) = 0$ for $t \le 0$, and $F(t) = 1$ for $t \ge 1$. Then $T \sim [0,1]$, and we have $B_{d_1}(t, \varepsilon) = \{\, s \in [0,1] \mid t - R_\varepsilon < s < t + R_\varepsilon \,\}$ for $t \in T$ and $\varepsilon > 0$, where $R_\varepsilon = 1/(\exp(1/\varepsilon^2) - 1)$. Let m be a probability measure on (T, d_1) for which the distribution function $H(t) = m([0,t])$ with $t \in [0,1]$ satisfies $H \in C^1[0,1]$. Then there exists $t_\varepsilon \in \,]t - R_\varepsilon, t + R_\varepsilon[$ such that $H(t + R_\varepsilon) - H(t - R_\varepsilon) = 2R_\varepsilon H'(t_\varepsilon)$ where $t \in \,]0,1[$ is given and fixed. Since $\psi^{-1}(y) = \sqrt{\log(1+y)}$, hence we get:

$$\int_0^{D_1} \psi^{-1}\Big(\frac{1}{m(B_{d_1}(t,\varepsilon))}\Big)\; d\varepsilon = \int_0^{D_1} \psi^{-1}\Big(\frac{1}{H(t+R_\varepsilon)-H(t-R_\varepsilon)}\Big)\; d\varepsilon$$

$$\sim \int_0^{D_1} \psi^{-1}\Big(\frac{1}{2R_\varepsilon H'(t_\varepsilon)}\Big)\; d\varepsilon = \int_0^1 \sqrt{\log\Big(1+\frac{1}{2H'(t_\varepsilon)R_\varepsilon}\Big)}\; d\varepsilon$$

$$= \int_0^1 \sqrt{\log\Big(1+\frac{\exp(1/\varepsilon^2)-1}{2H'(t_\varepsilon)}\Big)}\; d\varepsilon := \gamma\ .$$

Now, if $2H'(t_\varepsilon) \le 1$ then $\gamma \ge \int_0^1 \frac{1}{\varepsilon}\, d\varepsilon = +\infty$. If $2H'(t_\varepsilon) > 1$, then we have:

$$\gamma = \int_0^1 \sqrt{\log\Big(\frac{2H'(t_\varepsilon)+\exp(1/\varepsilon^2)-1}{2H'(t_\varepsilon)}\Big)}\; d\varepsilon \ge \int_0^1 \sqrt{\frac{1}{\varepsilon^2}-\log\big(2H'(t_\varepsilon)\big)}\; d\varepsilon = +\infty\ .$$

Thus, in any case we get $\gamma = +\infty$. Hence we find:

$$\int_0^{D_1} \psi^{-1}\Big(\frac{1}{m(B_{d_1}(t,\varepsilon))}\Big)\; d\varepsilon = +\infty\ .$$

Thus (2.178) fails for $n = 1$ and any probability measure m on (T, d_1) with distribution function from $C^1(\mathbf{R})$. The general case could be treated similarly.

2° Let $\psi(x) = \exp(x)-1$ for $x \ge 0$. Then $\|X_s^n - X_t^n\|_\psi = 1/(n\log(1+(2^n-1)/G(s,t))) := d_n(s,t)$. Let $F \sim U(0,1)$ and m with H be as in 1° above. Put $R_\varepsilon = (2^n-1)/(\exp(1/n\varepsilon)-1)$, then by the same arguments as in 1° above we obtain:

$$\int_0^{D_n} \psi^{-1}\Big(\frac{1}{m(B_{d_n}(t,\varepsilon))}\Big)\; d\varepsilon = \int_0^{D_n} \psi^{-1}\Big(\frac{1}{H(t+R_\varepsilon)-H(t-R_\varepsilon)}\Big)\; d\varepsilon$$

$$\sim \int_0^{D_n} \psi^{-1}\Big(\frac{1}{2H'(t_\varepsilon)R_\varepsilon}\Big)\; d\varepsilon = \int_0^{D_n} \log\Big(1+\frac{\exp(1/n\varepsilon)-1}{2\,(2^n-1)\,H'(t_\varepsilon)}\Big)\; d\varepsilon = +\infty$$

for all $n \ge 1$ and all $t \in \,]0,1[\, \sim T$. Thus (2.178) fails for all $n \ge 1$ and any probability measure m on (T, d_n) with distribution function from $C^1(\mathbf{R})$. The general case could be treated similarly.

3. The Vapnik-Chervonenkis Approach

In this chapter we present the Vapnik-Chervonenkis approach towards uniform laws of large numbers which appeared in [89] and [90], and its extension to completely regular dynamical systems which is from [68]. The first section presents the classical subgaussian inequality for Rademacher series. It is a cornerstone in the sufficiency part of the VC theorem in the fourth section (obtained through the intervention of Rademacher randomization given in the third section). The second section presents Sudakov's minoration [83] (via Slepian's lemma [78]) which is a cornerstone in the necessity part of the VC theorem in the fourth section (obtained through the intervention of Gaussian randomization given in the third section). The fifth section is devoted to the VC classes of sets, which provide the best known examples satisfying the VC theorem. The sixth section presents a lemma of Eberlein (see [23]) which is at the basis of the extension of the VC theorem to completely regular dynamical systems. Such a theorem is given in the seventh section and is taken from [68]. In the eighth section we present its extension to semi-flows and (non-linear) operators (from [68] as well). The ninth section deals with a special problem in this context on uniformity over factorizations (this is also taken from [68]). The tenth section is reserved for examples and complements. Counter-examples of Weber and Nobel given there indicate that the VC random entropy numbers approach is essentially linked with dynamical systems which posses a mixing property (thus being far from the general stationary ergodic case for example). Completely regular dynamical systems seem to be somewhere around the border of its successful applicability.

3.1 The subgaussian inequality for Rademacher series

The sufficiency part of the VC theorem (in Section 3.4 below) follows essentially from the next classic inequality (after performing a Rademacher randomization as explained in Section 3.3 below). This inequality is known to be extremely useful in many other contexts as well. We recall that a sequence of random variables $\{\varepsilon_j\}_{j\geq 1}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ is said to be a *Rademacher sequence*, if ε_j's are independent and $P\{\varepsilon_j = \pm 1\} = 1/2$ for $j \geq 1$.

Theorem 3.1 (The subgaussian inequality)

Let $\{\varepsilon_j\}_{j\geq 1}$ be a Rademacher sequence, and let $\{a_j\}_{j\geq 1}$ be a sequence of real numbers. Denote $S_n = \sum_{j=1}^n a_j \varepsilon_j$ and $A_n = \sum_{j=1}^n |a_j|^2$ for $n \geq 1$. Then we have:

$$P\{\, |S_n| > t \,\} \;\leq\; 2\exp\Big(-\frac{t^2}{2A_n}\Big) \tag{3.1}$$

for all $t > 0$ and all $n \geq 1$.

Proof. Let $t > 0$ and $n \geq 1$ be given and fixed. Since $\cosh(x) \leq \exp(x^2/2)$ for $x \in \mathbf{R}$, then by symmetry, Markov's inequality and independence we get:

$$P\{\, |S_n| > t \,\} = 2P\{\, S_n > t \,\} = 2P\{\, \exp(\lambda S_n) > \exp(\lambda t) \,\}$$

$$\leq 2\exp(-\lambda t)\; E\exp(\lambda S_n) = 2\exp(-\lambda t)\; E\prod_{j=1}^n \exp(\lambda a_j \varepsilon_j)$$

$$= 2\exp(-\lambda t)\; \prod_{j=1}^n E\exp(\lambda a_j \varepsilon_j) = 2\exp(-\lambda t)\; \prod_{j=1}^n \cosh(\lambda a_j)$$

$$\leq 2\exp(-\lambda t)\; \prod_{j=1}^n \exp\big(\lambda^2 |a_j|^2/2\big) = 2\exp\big(\lambda^2 A_n/2 - \lambda t\big)$$

for all $\lambda > 0$. Minimizing the right-hand side over $\lambda > 0$, we obtain $\lambda = t/A_n$. Inserting this λ into the inequality just obtained, we get (3.1). The proof is complete. $\square$

3.2 Slepian's lemma and Sudakov's minoration

The necessity part of the VC theorem (in Section 3.4 below) relies essentially upon *Sudakov's minoration* [83] (after performing a Gaussian randomization as explained in Section 3.3 below). The result is presented in Theorem 3.5 below, and is based on a comparison fact from Corollary 3.3 (which in turn relies upon *Slepian's lemma* [78] from Theorem 3.2) and a classical Gaussian estimate from Lemma 3.4. Our exposition of these results follows the approach taken in [54].

We clarify that when we speak of a *Gaussian* random variable, we always mean a *centered* Gaussian variable (its expectation is zero). A Gaussian random variable is called *standard*, if its variance equals one. A random vector $X = (X_1, \ldots, X_n)$ with values in $\mathbf{R}^n$ is said to be Gaussian, if $\sum_{i=1}^n \alpha_i X_i$ is a Gaussian random variable for all $\alpha_1, \ldots, \alpha_n \in \mathbf{R}$. A real valued random process $X = (X_t)_{t \in T}$ indexed by a set T is said to be Gaussian, if $\sum_{i=1}^n \alpha_i X_{t_i}$ is a Gaussian random variable for all $\alpha_1, \ldots, \alpha_n \in \mathbf{R}$, all $t_1, \ldots, t_n \in T$, and all $n \geq 1$.

Theorem 3.2 (Slepian's lemma 1962)
Let $X = (X_1, \ldots, X_n)$ *and* $Y = (Y_1, \ldots, Y_n)$ *be Gaussian random vectors satisfying:*

$$E(X_i X_j) \leq E(Y_i Y_j) \quad \forall\, i \neq j \in \{1, \ldots, n\} \tag{3.2}$$

$$E(X_i^2) = E(Y_i^2) \quad \forall\, i = 1, \ldots, n \, . \tag{3.3}$$

Then the following inequality is satisfied:

$$E\Big(\max_{1 \leq i \leq n} Y_i\Big) \leq E\Big(\max_{1 \leq i \leq n} X_i\Big) \, . \tag{3.4}$$

Proof. We may and do assume that X and Y are independent. Define the process in $\mathbf{R}^n$ by:

$$Z(t) \, = \, (\sqrt{1-t}\,)\, X + (\sqrt{t}\,)\, Y$$

for $t \in [0,1]$. Put $M(x) = \max\{x_1, \ldots, x_n\}$ for $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. Consider the map:

$$G(t) \, = \, E\Big(M\big(Z(t)\big)\Big)$$

for $t \in [0,1]$. The proof consists of showing that G is decreasing on $[0,1]$. This fact implies $G(1) \leq G(0)$, which is precisely (3.4).

To show that G is decreasing, we compute its derivative:

$$G'(t) \, = \, \sum_{i=1}^n E\Big(\frac{\partial M}{\partial x_i}\big(Z(t)\big) \cdot Z_i'(t)\Big)$$

for all $t \in [0,1]$. (It should be noted that $\partial M/\partial x_i$ is well-defined outside the union of the hyperplanes $\{\, x \in \mathbf{R}^n \mid x_j = x_k \,\}$ in $\mathbf{R}^n$ for $1 \leq i, j, k \leq n$, which is a nullset with respect to the Gaussian measure.) In order to show that $G'(t) \leq 0$ for $t \in [0,1]$ (which implies that G is decreasing), we will establish the inequality:

$$E\Big(\frac{\partial M}{\partial x_i}\big(Z(t)\big)\cdot Z_i'(t)\Big) \;\leq\; 0 \tag{3.5}$$

for all $1 \leq i \leq n$ and all $t \in [0,1]$. For this, let us fix $1 \leq i \leq n$ and $t \in [0,1]$, let us put $\gamma_j^* = E\big(Z_j(t)Z_i'(t)\big)/E\big(Z_i'(t)\big)^2$, and let the vector $W = (W_1, \ldots, W_n)$ be given by the equation:

$$Z_j(t) \;=\; \gamma_j^* \, Z_i'(t) + W_j$$

for all $1 \leq j \leq n$. Now, first note that $E\big(Z_j(t)Z_i'(t)\big) = E(Y_iY_j - X_iX_j)/2$ which is by (3.2) and (3.3) non-negative. Thus $\gamma_j^* \geq 0$ for all $1 \leq j \leq n$, and $\gamma_i^* = 0$. Next, let us consider the map:

$$H(\gamma) = E\bigg(\frac{\partial M}{\partial x_i}\Big(\gamma_1 Z_i'(t) + W_1, \;\ldots\;, \gamma_n Z_i'(t) + W_n\Big)\cdot Z_i'(t)\bigg)$$

for $\gamma = (\gamma_1, \ldots, \gamma_n) \in \mathbf{R}_+^n$. Since $\partial M/\partial x_i$ is decreasing in each component different from the i-th one, it is easily verified that H is so as well. Moreover, by the choice of γ_j's we may easily verify that $E(W_j Z_i'(t)) = 0$ for all $1 \leq j \leq n$. Since W_j and $Z_i'(t)$ are Gaussian variables, then they are independent for all $1 \leq j \leq n$. Thus W and $Z_i'(t)$ are independent, and therefore $E\big(\,(\partial M/\partial x_i)(W)\cdot Z_i'(t)\,\big) = 0$. This shows $H(0) = 0$, which together with the fact that H is decreasing in each component different from the i-th one implies $H(\gamma_1^*, \ldots, \gamma_n^*) \leq 0$. This establishes (3.5), and the proof is complete. □

Corollary 3.3
Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ be Gaussian random vectors satisfying:

$$E|Y_i - Y_j|^2 \;\leq\; E|X_i - X_j|^2 \tag{3.6}$$

for all $1 \leq i,j \leq n$. Then the following inequality is satisfied:

$$E\Big(\max_{1\leq i\leq n} Y_i\Big) \leq 2E\Big(\max_{1\leq i\leq n} X_i\Big)\;. \tag{3.7}$$

Proof. It is no restriction to assume that $X_1 = Y_1 = 0$. (Otherwise we could replace X and Y by $(X_1 - X_1, \ldots, X_n - X_1)$ and $(Y_1 - Y_1, \ldots, Y_n - Y_1)$ respectively.) Put $\sigma^2 = \max_{1\leq i\leq n} E|X_i|^2$, and consider the Gaussian random variables:

$$\hat{X}_i = X_i + \sqrt{\big(\sigma^2 + E|Y_i|^2 - E|X_i|^2\big)}\cdot g \quad \text{and} \quad \hat{Y}_i = Y_i + \sigma\cdot g$$

for $1 \leq i \leq n$, where g is a standard Gaussian variable which is independent from X and Y .

We verify that the Gaussian vectors $\hat{X} = (\hat{X}_1, \ldots, \hat{X}_n)$ and $\hat{Y} = (\hat{Y}_1, \ldots, \hat{Y}_n)$ satisfy (3.2) and (3.3) from Theorem 3.2. For this, first note that $E|\hat{X}_i|^2 = E|\hat{Y}_i|^2 = \sigma^2 + E|Y_i|^2$ for all $1 \leq i \leq n$. Thus (3.3) is satisfied. Next, note that by (3.6) we get:

$$E|\hat{Y}_i - \hat{Y}_j|^2 = E|Y_i - Y_j|^2 \leq E|X_i - X_j|^2 \leq E|\hat{X}_i - \hat{X}_j|^2$$

for all $1 \leq i,j \leq n$. Hence $E(\hat{X}_i\hat{X}_j) \leq E(\hat{Y}_i\hat{Y}_j)$ for all $i \neq j \in \{1, \ldots, n\}$. Thus (3.2) is satisfied. Applying Theorem 3.2 we obtain:

$$E\Big(\max_{1\leq i\leq n} \hat{Y}_i\Big) \leq E\Big(\max_{1\leq i\leq n} \hat{X}_i\Big)\;. \tag{3.8}$$

It is clear that we have:

$$(3.9) \qquad E\Big(\max_{1\le i\le n} \hat{Y}_i\Big) = E\Big(\max_{1\le i\le n} Y_i\Big) \ .$$

Moreover, since by assumption $X_1 = Y_1 = 0$, then by (3.6) we have $E|Y_i|^2 \le E|X_i|^2$ for all $1 \le i \le n$. Hence we get:

$$(3.10) \qquad E\Big(\max_{1\le i\le n} \hat{X}_i\Big) \le E\Big(\max_{1\le i\le n} X_i\Big) + \sigma E(g^+) \ .$$

Finally, since $X_1 = 0$ we easily conclude by symmetry:

$$(3.11) \qquad \sigma = \max_{1\le i\le n} \sqrt{E|X_i|^2} = \frac{1}{E|g|} \max_{1\le i\le n} E|X_i| \le \frac{1}{E|g|} E\Big(\max_{1\le i,j\le n} |X_i - X_j|\Big)$$
$$= \frac{2}{E|g|} E\Big(\max_{1\le i\le n} X_i\Big) = \frac{1}{E(g^+)} E\Big(\max_{1\le i\le n} X_i\Big) \ .$$

Now (3.7) follows from (3.8)-(3.11), and the proof is complete. □

Lemma 3.4
Let $X = (X_1, \dots, X_n)$ be a Gaussian random vector, and let $g_1, \dots, g_n$ be independent standard Gaussian random variables. Then the inequalities are satisfied:

$$(3.12) \qquad E\Big(\max_{1\le i\le n} |X_i|\Big) \le 3\sqrt{\log n} \max_{1\le i\le n} \sqrt{E|X_i|^2}$$

$$(3.13) \qquad \frac{1}{K} \sqrt{\log n} \le E\Big(\max_{1\le i\le n} g_i\Big) \le K \cdot \sqrt{\log n}$$

with a constant $K > 0$ not dependent on $n > 1$.

Proof. (3.12): It is no restriction to assume that $\max_{1\le i\le n} \sqrt{E|X_i|^2} = 1$. From integration by parts we easily find:

$$(3.14) \qquad E\Big(\max_{1\le i\le n} |X_i|\Big) = \int_0^\infty P\Big\{ \max_{1\le i\le n} |X_i| > t \Big\} \, dt \le \delta + n \int_\delta^\infty P\{ |g| > t \} \, dt$$

where g is a standard Gaussian random variable. We recall the classic estimate:

$$(3.15) \qquad P\{ g > t \} \le \tfrac{1}{2} \exp(-t^2/2)$$

being valid for all $t > 0$. Applying (3.15) twice in (3.14) we get:

$$E\Big(\max_{1\le i\le n} |X_i|\Big) \le \delta + n \sqrt{\pi/2} \exp(-\delta^2/2)$$

which is valid for all $\delta > 0$. Taking $\delta = \sqrt{2\log n}$ we obtain:

$$E\Big(\max_{1\le i\le n} |X_i|\Big) \le \sqrt{2\log n} + \sqrt{\pi/2} \le 3\sqrt{\log n}$$

whenever $n \ge 2$. This completes the proof of (3.12).

(3.13): We only prove the first inequality, since the second inequality follows straightforwardly from (3.12). Integration by parts, independence and identical distribution yield:

$$(3.16) \qquad E\Big(\max_{1\le i\le n}|g_i|\Big) \ge \int_0^\Delta \big(1-(1-P\{|g|>t\})^n\big)\ dt \ge \Delta\big(1-(1-P\{|g|>\Delta\})^n\big)$$

for all $\Delta > 0$, where g is a standard Gaussian random variable. Further, we find:

$$P\{|g|>\Delta\} = \frac{2}{\sqrt{2\pi}}\int_\Delta^\infty \exp(-t^2/2)\ dt \ge \frac{2}{\sqrt{2\pi}}\exp(-(\Delta+1)^2/2)$$

for all $\Delta > 0$. Taking $\Delta = \sqrt{\log n}$ with $n \ge n_0$ large enough so that $P\{|g|>\Delta\} \ge 1/n$, from (3.16) we obtain:

$$(3.17) \qquad E\Big(\max_{1\le i\le n}|g_i|\Big) \ge \sqrt{\log n}\,\Big(1-\big(1-\tfrac{1}{n}\big)^n\Big) \le \sqrt{\log n}\,\Big(1-\tfrac{1}{e}\Big)\ .$$

Now, by the triangle inequality and symmetric distribution we find:

$$(3.18) \qquad E\Big(\max_{1\le i\le n}|g_i|\Big) \le E\Big(\max_{1\le i,j\le n}|g_i-g_j|\Big) + E|g| = 2E\Big(\max_{1\le i\le n} g_i\Big) + \sqrt{2/\pi}$$

for $n \ge 1$. Combining (3.17) and (3.18) we get $E\big(\max_{1\le i\le n} g_i\big) \ge 1/2\,\big(\sqrt{\log n}\,(1-\frac{1}{e})-\sqrt{2/\pi}\big)$ for all $n \ge n_0$. Letting K to be large enough, we obtain (3.13) and complete the proof. $\square$

Given a pseudo-metric space (T,d) , we denote by $N(T,d;\varepsilon)$ the *entropy numbers* associated with the pseudo-metric d on the set T for $\varepsilon > 0$. We recall that $N(T,d;\varepsilon)$ is the smallest number of open balls of radius $\varepsilon > 0$ in the pseudo-metric d needed to cover the set T (see the Definition in the beginning of Chapter 2).

Theorem 3.5 (Sudakov's minoration 1971)
Let $X=(X_t)_{t\in T}$ *be a Gaussian process indexed by a set* T *with its intrinsic pseudo-metric* $d_X(s,t) = \|X_s-X_t\|_2$ *for* $s,t\in T$ *. Then there exists a constant* $C > 0$ *such that:*

$$(3.19) \qquad \varepsilon\sqrt{\log N\big(T,d_X;\varepsilon\big)} \le C\,E\Big(\sup_{t\in T} X_t\Big)$$

being valid for all $\varepsilon > 0$ *.*

Proof. Given $\varepsilon > 0$, let $N = N(T,d_X;\varepsilon)$. Then there exists a finite set $F\subset T$ with $card\,(F) = N$ such that $d_X(s,t) > \varepsilon$ for all $s,t\in F$, $s\ne t$. Let $G=(g_t)_{t\in F}$ be a family of independent standard Gaussian random variables indexed by F , and let us put $\hat{g}_t = (\varepsilon/\sqrt{2})\,g_t$ for $t\in F$. Then we have:

$$\|\hat{g}_s-\hat{g}_t\|_2 = (\varepsilon/\sqrt{2})\,\|g_s-g_t\|_2 = \varepsilon < d_X(s,t) = \|X_s-X_t\|_2$$

for all $s,t\in F$. Thus by Corollary 3.3 we get:

$$E\Big(\sup_{t\in F}\hat{g}_t\Big) \le 2E\Big(\sup_{t\in F} X_t\Big)\ .$$

Finally, by Lemma 3.4 we have:

$$E\Big(\sup_{t\in F}\hat{g}_t\Big) \ge \frac{\varepsilon}{\sqrt{2}}\,\frac{1}{K}\sqrt{\log\big(card\,(F)\big)} = \frac{\varepsilon}{K\sqrt{2}}\,\sqrt{\log N}\ .$$

The last two inequalities complete the proof. $\square$

Remark 3.6

The proof shows that the supremum appearing in (3.19) should be understood as follows:

$$E\Big(\sup_{t\in T} X_t\Big) = \sup\Big\{\, E\Big(\sup_{t\in F} X_t\Big) \;\Big|\; F\subset T \;\text{ is finite}\,\Big\}\ .$$

Thus, in general, we could replace $E\big(\sup_{t\in T} X_t\big)$ in (3.19) by $E^*\big(\sup_{t\in T} X_t\big)$.

Remark 3.7

The proof shows that the constant $C>0$ appearing in (3.19) does not depend on the Gaussian process $X=(X_t)_{t\in T}$ itself. Actually, we have $C=2\sqrt{2}\,K$, where K is the constant appearing in the first inequality in (3.13).

3.3 Rademacher and Gaussian randomization

Our main aim in this section is to describe the procedures of Rademacher and Gaussian randomization in the context of the uniform law of large numbers. These procedures play the most important role in the approach of Vapnik and Chervonenkis (at least from the modern point of view of the theory today). The exposition is inspired by [32], and the credits for the rather classical facts given below may be found there.

1. Throughout we consider a sequence of independent and identically distributed random variables $\xi = \{\, \xi_j \mid j\geq 1\,\}$ defined on the probability space $(\Omega,\mathcal{F},P)$ with values in the measurable space $(S,\mathcal{A})$ and a common distribution law π . We suppose that a set T is given, as well as a map $f: S\times T\to \mathbf{R}$ such that $s\mapsto f(s,t)$ is π-measurable for all $t\in T$. We denote $M(t)=\int_S f(s,t)\,\pi(ds)$ for $t\in T$. Through the whole section we assume that the structure of f and T is flexible enough, such that supremums over T which involve f define measurable maps. (We recall that this assumption is not restrictive and can be supported in quite a general setting by using theory of analytic spaces as explained in Paragraph 5 of Introduction.)

In order to perform the Rademacher randomization we suppose that a Rademacher sequence $\varepsilon=\{\varepsilon_j \mid j\geq 1\}$ is given (see Section 3.1), which is defined on the probability space $(\Omega_\varepsilon,\mathcal{F}_\varepsilon,P_\varepsilon)$. Whenever the sequences ξ and ε appear together, we assume that they are both defined on the product probability space $(\Omega\otimes\Omega_\varepsilon,\mathcal{F}\otimes\mathcal{F}_\varepsilon,P\otimes P_\varepsilon)$, and thus may be understood mutually independent. Similarly, in order to perform the Gaussian randomization we suppose that a standard Gaussian sequence $g=\{g_j\mid j\geq 1\}$ is given, meaning that g_j for $j\geq 1$ are independent standard Gaussian variables (see Section 3.2), which is defined on the probability space $(\Omega_g,\mathcal{F}_g,P_g)$. Also, whenever the sequences ξ and g appear together, we assume that they are both defined on the product probability space $(\Omega\otimes\Omega_g,\mathcal{F}\otimes\mathcal{F}_g,P\otimes P_g)$, and thus may be understood mutually independent. Finally, whenever the three sequences ξ , ε and g appear together, we assume that they are all defined on the product probability space $(\Omega\otimes\Omega_\varepsilon\otimes\Omega_g,\mathcal{F}\otimes\mathcal{F}_\varepsilon\otimes\mathcal{F}_g,P\otimes P_\varepsilon\otimes P_g)$, and thus may be understood mutually independent.

2. We are primarily interested in the asymptotic behaviour of the expression:

$$\sup_{t\in T}\,\Big|\,\frac{1}{n}\sum_{j=1}^{n} f(\xi_j,t)\,-M(t)\,\Big| \tag{3.20}$$

as $n\to\infty$. The *Rademacher randomization* consists of forming (by ε randomized) expression:

$$(3.21) \qquad \sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot f(\xi_j, t) \right|$$

and comparing it with the expression in (3.20) as $n \to \infty$. The *Gaussian randomization* consists of forming (by g randomized) expression:

$$(3.22) \qquad \sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^{n} g_j \cdot f(\xi_j, t) \right|$$

and comparing it with the expression in (3.20) as $n \to \infty$. The main purpose of this section is to show that the expressions in (3.20), (3.21) and (3.22) have the same asymptotic behaviour for $n \to \infty$. The main gain of this relies upon the facts that the process $(\omega', t) \mapsto \sum_{j=1}^{n} \varepsilon_j(\omega') f(\xi_j, t)$ is subgaussian (recall Theorem 3.1) and the process $(\omega'', t) \mapsto \sum_{j=1}^{n} g_j(\omega'') f(\xi_j, t)$ is Gaussian (recall Theorem 3.5). Thus, by Rademacher and Gaussian randomization we imbed the problem of the asymptotic behaviour of the expression in (3.20) into theory of Gaussian (subgaussian) processes. The result just indicated may be stated more precisely as follows.

Theorem 3.8

Let $\xi = \{\xi_j \mid j \geq 1\}$ be a sequence of independent and identically distributed random variables with values in the measurable space $(S, \mathcal{A})$ and a common distribution law π , let $\varepsilon = \{\varepsilon_j \mid j \geq 1\}$ be a Rademacher sequence, and let $g = \{ g_j \mid j \geq 1 \}$ be a standard Gaussian sequence. Let T be a set, and let $f : S \times T \to \mathbf{R}$ be a function such that $s \mapsto f(s, t)$ is π-measurable for all $t \in T$, and such that all supremums over T taken below are measurable (possibly with respect to the completed σ-algebras). Suppose moreover that $\| f \|_T \in L^1(\pi)$. Then the following three convergence statements are equivalent:

$$(3.23) \qquad E\left(\sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j, t) - M(t) \right| \right) \longrightarrow 0$$

$$(3.24) \qquad E\left(\sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot f(\xi_j, t) \right| \right) \longrightarrow 0$$

$$(3.25) \qquad E\left(\sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^{n} g_j \cdot f(\xi_j, t) \right| \right) \longrightarrow 0$$

as $n \to \infty$. (We clarify that the symbol E in (3.23) denotes the P-integral, the symbol E in (3.24) denotes the $P \otimes P_\varepsilon$-integral, and the symbol E in (3.25) denotes the $P \otimes P_g$-integral.) Moreover, the mean convergence in either of (3.23)-(3.25) may be equivalently replaced by either almost sure convergence or convergence in probability.

Proof. We first remark that the last statement of the theorem follows from Corollary 2.8. In this context we clarify that for (3.24) one should consider $S' = S \times \{-1, 1\}$ and the map $F : S' \times T \to \mathbf{R}$ defined by $F((s, s'), t) = s' f(s, t)$, while for (3.25) one should consider $S'' = S \times \mathbf{R}$ and the map $G : S'' \times T \to \mathbf{R}$ defined by $G((s, s''), t) = s'' f(s, t)$. Then the claim is easily verified.

We turn to the equivalence between (3.23)-(3.25). First we consider the case (3.23)-(3.24).

Lemma 3.9

Under the hypotheses in Theorem 3.8 we have:

$$E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n f(\xi_j,t)-M(t)\Big|\Big)\le 2E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\cdot f(\xi_j,t)\Big|\Big) \tag{3.26}$$

$$E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\cdot f(\xi_j,t)\Big|\Big)\le 4\max_{1\le k\le n}E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^k f(\xi_j,t)-M(t)\Big|\Big)+ \tag{3.27}$$

$$+\,\|M\|_T\cdot E\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\Big|$$

for all $n\ge 1$.

Proof of Lemma 3.9. (3.26): Let $\xi'=\{\,\xi'_j\mid j\ge 1\,\}$ be an independent copy of $\xi=\{\,\xi_j\mid j\ge 1\,\}$ defined on $(\Omega',\mathcal{F}',P')$. Then we have:

$$E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n f(\xi_j,t)-M(t)\Big|\Big)=E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n f(\xi_j,t)-E'f(\xi'_j,t)\Big|\Big)$$

$$\le EE'\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n f(\xi_j,t)-f(\xi'_j,t)\Big|\Big)=EE'\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\big(f(\xi_j,t)-f(\xi'_j,t)\big)\Big|\Big)$$

$$\le 2E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\cdot f(\xi_j,t)\Big|\Big)$$

for any given and fixed (numbers) $\varepsilon_1,\dots,\varepsilon_n$ with $n\ge 1$. The inequality follows by taking the P_ε-integral on both sides.

(3.27): Replacing $f(s,t)$ by $f(s,t)-M(t)$, and using $|x-y|\ge|x|-|y|$, we may and do assume that $M(t)=0$ for all $t\in T$. Thus the inequality to be proved is as follows:

$$E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\cdot f(\xi_j,t)\Big|\Big)\le 4\max_{1\le k\le n}E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^k f(\xi_j,t)\Big|\Big) \tag{3.28}$$

for $n\ge 1$ being fixed. For this, note that we have:

$$P\otimes P_\varepsilon\Big\{\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\cdot f(\xi_j,t)\Big|>x\Big\}$$

$$=\frac{1}{2^n}\sum_{(\varepsilon_1,\dots,\varepsilon_n)\in\{-1,1\}^n}P\Big\{\sup_{t\in T}\Big|\frac{1}{n}\sum_{\varepsilon_j=1}f(\xi_j,t)-\frac{1}{n}\sum_{\varepsilon_j=-1}f(\xi_j,t)\Big|>x\Big\}$$

$$\le\frac{1}{2^n}\sum_{(\varepsilon_1,\dots,\varepsilon_n)\in\{-1,1\}^n}\Big(P\Big\{\sup_{t\in T}\Big|\frac{1}{n}\sum_{\varepsilon_j=1}f(\xi_j,t)\Big|>\frac{x}{2}\Big\}+P\Big\{\sup_{t\in T}\Big|\frac{1}{n}\sum_{\varepsilon_j=-1}f(\xi_j,t)\Big|>\frac{x}{2}\Big\}\Big)$$

for all $x>0$. Taking the integral on both sides, we obtain by integration by parts:

$$E\Big(\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\cdot f(\xi_j,t)\Big|\Big)=\int_0^\infty P\otimes P_\varepsilon\Big\{\sup_{t\in T}\Big|\frac{1}{n}\sum_{j=1}^n \varepsilon_j\cdot f(\xi_j,t)\Big|>x\Big\}\,dx\ \le$$

$$\leq \frac{1}{2^n} \sum_{(\varepsilon_1,\dots,\varepsilon_n)\in\{-1,1\}^n} 2 \max_{1\leq k\leq n} \int_0^\infty P\Big\{ \sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^k f(\xi_j,t) \Big| > \frac{x}{2} \Big\}\, dx$$

$$= 4 \max_{1\leq k\leq n} \int_0^\infty P\Big\{ \sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^k f(\xi_j,t) \Big| > x \Big\}\, dx \;=\; 4 \max_{1\leq k\leq n} E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^k f(\xi_j,t) \Big| \Big)\ .$$

Thus (3.28) is valid, and (3.27) follows as well. The proof of Lemma 3.9 is complete. □

3. We now show how to use Lemma 3.9 to establish the equivalence between (3.23) and (3.24). First, if (3.24) is valid, then we get (3.23) straightforwardly from (3.26). Second, suppose that (3.23) is satisfied. Since $E\,|\,(1/n)\sum_{j=1}^n \varepsilon_j\,| = o(1)$ for $n \to \infty$, by (3.27) it is enough to show that:

$$\max_{1\leq k\leq n} E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^k f(\xi_j,t) - M(t) \Big| \Big) \longrightarrow 0 \tag{3.29}$$

as $n \to \infty$. For this, note that we have:

$$\max_{1\leq k\leq n} E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^k f(\xi_j,t) - M(t) \Big| \Big)$$

$$\leq \max_{1\leq k\leq n_0} \frac{k}{n} E\Big(\sup_{t\in T} \Big| \frac{1}{k} \sum_{j=1}^k f(\xi_j,t) - M(t) \Big| \Big) + \max_{n_0<k\leq n} \frac{k}{n} E\Big(\sup_{t\in T} \Big| \frac{1}{k} \sum_{j=1}^k f(\xi_j,t) - M(t) \Big| \Big)$$

$$\leq \frac{n_0}{n} E\Big(\sup_{t\in T} \Big| f(\xi_1,t) - M(t) \Big| \Big) + \max_{n_0<k\leq n} E\Big(\sup_{t\in T} \Big| \frac{1}{k} \sum_{j=1}^k f(\xi_j,t) - M(t) \Big| \Big)$$

for all $n \geq n_0 \geq 1$. Now, given $\varepsilon > 0$, take $n_0 \geq 1$ such that the last expression is less than ε . Then take the limit superior on both sides as $n \to \infty$. Since $(n_0/n)E\big(\sup_{t\in T} |f(\xi_1,t) - M(t)|\big) \to 0$ as $n \to \infty$, we obtain (3.29). This completes the proof of the equivalence (3.23)-(3.24).

Next we turn to the equivalence between (3.24) and (3.25).

Lemma 3.10

Under the hypotheses in Theorem 3.8 we have:

$$E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot f(\xi_j,t) \Big| \Big) \leq \sqrt{\frac{\pi}{2}}\, E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^n g_j \cdot f(\xi_j,t) \Big| \Big) \tag{3.30}$$

$$E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=n_0}^n g_j \cdot f(\xi_j,t) \Big| \Big) \leq \sqrt{\frac{2}{\pi}} \max_{n_0\leq k\leq n} E\Big(\sup_{t\in T} \Big| \frac{1}{k} \sum_{j=n_0}^k \varepsilon_j \cdot f(\xi_j,t) \Big| \Big) \tag{3.31}$$

for all $n \geq n_0 \geq 1$.

Proof of Lemma 3.10. (3.30): We have:

$$E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^n g_j \cdot f(\xi_j,t) \Big| \Big) = E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot |g_j| \cdot f(\xi_j,t) \Big| \Big)$$

$$= \int_\Omega \int_{\Omega_\varepsilon} \int_{\Omega_g} \Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot |g_j| \cdot f(\xi_j,t) \Big| \Big)\, dP\, dP_\varepsilon\, dP_g \geq$$

$$\geq \int_{\Omega}\int_{\Omega_\varepsilon}\left(\sup_{t\in T}\left|\int_{\Omega_g}\left(\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\cdot|g_j|\cdot f(\xi_j,t)\right)dP_g\right|\right)dP\,dP_\varepsilon$$
$$= \int_{\Omega}\int_{\Omega_\varepsilon}\left(\sup_{t\in T}\left|\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\cdot E|g_1|\cdot f(\xi_j,t)\right|\right)dP\,dP_\varepsilon = \sqrt{\frac{2}{\pi}}E\left(\sup_{t\in T}\left|\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\cdot f(\xi_j,t)\right|\right)$$

for all $n \geq 1$. This establishes (3.30) and completes the first part of the proof.

(3.31): Let $n \geq n_0 \geq 1$ be given and fixed, and let $|g_1|^* \geq |g_2|^* \geq \ldots \geq |g_n|^* \geq |g_{n+1}|^* := 0$ be the non-increasing rearrangement of the sequence $\big(|g_i|\big)_{1\leq i\leq n}$. Then we have:

$$E\left(\sup_{t\in T}\left|\sum_{j=n_0}^{n}g_j\cdot f(\xi_j,t)\right|\right)$$
$$= E\left(\sup_{t\in T}\left|\sum_{j=n_0}^{n}|g_j|\cdot\varepsilon_j\cdot f(\xi_j,t)\right|\right) = E\left(\sup_{t\in T}\left|\sum_{j=n_0}^{n}|g_j|^*\cdot\varepsilon_j\cdot f(\xi_j,t)\right|\right)$$
$$= E\left(\sup_{t\in T}\left|\sum_{k=n_0}^{n}\big(|g_k|^*-|g_{k+1}|^*\big)\cdot\sum_{j=n_0}^{k}\varepsilon_j\cdot f(\xi_j,t)\right|\right)$$
$$\leq E\left|\sum_{k=n_0}^{n}k\big(|g_k|^*-|g_{k+1}|^*\big)\right|\cdot\max_{n_0\leq k\leq n}E\left(\sup_{t\in T}\left|\frac{1}{k}\sum_{j=n_0}^{k}\varepsilon_j\cdot f(\xi_j,t)\right|\right)$$
$$\leq E\left|\sum_{k=1}^{n}k\big(|g_k|^*-|g_{k+1}|^*\big)\right|\cdot\max_{n_0\leq k\leq n}E\left(\sup_{t\in T}\left|\frac{1}{k}\sum_{j=n_0}^{k}\varepsilon_j\cdot f(\xi_j,t)\right|\right)$$
$$= E\left(\sum_{k=1}^{n}|g_k|^*\right)\cdot\max_{n_0\leq k\leq n}E\left(\sup_{t\in T}\left|\frac{1}{k}\sum_{j=n_0}^{k}\varepsilon_j\cdot f(\xi_j,t)\right|\right)$$
$$= nE|g_1|\cdot\max_{n_0\leq k\leq n}E\left(\sup_{t\in T}\left|\frac{1}{k}\sum_{j=n_0}^{k}\varepsilon_j\cdot f(\xi_j,t)\right|\right).$$

Since $E|g_1| = \sqrt{2/\pi}$, this finishes the proof of (3.31). The proof of Lemma 3.10 is complete. $\square$

4. We now show how to use Lemma 3.10 to establish the equivalence between (3.24) and (3.25). First, if (3.25) is valid, then we get (3.24) straightforwardly from (3.30). Second, suppose that (3.24) is satisfied. Then by (3.31) we find:

$$E\left(\sup_{t\in T}\left|\frac{1}{n}\sum_{j=1}^{n}g_j\cdot f(\xi_j,t)\right|\right)$$
$$\leq E\left(\sup_{t\in T}\left|\frac{1}{n}\sum_{j=1}^{n_0-1}g_j\cdot f(\xi_j,t)\right|\right) + \sqrt{\frac{2}{\pi}}\max_{n_0\leq k\leq n}E\left(\sup_{t\in T}\left|\frac{1}{k}\sum_{j=n_0}^{k}\varepsilon_j\cdot f(\xi_j,t)\right|\right)$$
$$\leq \frac{(n_0-1)}{n}\cdot E|g_1|\cdot E\Big(\sup_{t\in T}\big|f(\xi_1,t)\big|\Big) + \sqrt{\frac{2}{\pi}}\max_{n_0\leq k\leq n}E\left(\sup_{t\in T}\left|\frac{1}{k}\sum_{j=1}^{k}\varepsilon_j\cdot f(\xi_j,t)\right|\right)+$$

$$+\sqrt{\frac{2}{\pi}}\max_{n_0\le k\le n} E\Big(\sup_{t\in T}\Big|\frac{1}{k}\sum_{j=1}^{n_0-1}\varepsilon_j\cdot f(\xi_j,t)\Big|\Big)$$

$$\le \frac{(n_0-1)}{n}\cdot\sqrt{\frac{2}{\pi}}\cdot E\Big(\sup_{t\in T}|f(\xi_1,t)|\Big)+\sqrt{\frac{2}{\pi}}\max_{n_0\le k\le n} E\Big(\sup_{t\in T}\Big|\frac{1}{k}\sum_{j=1}^{k}\varepsilon_j\cdot f(\xi_j,t)\Big|\Big)$$

$$+\sqrt{\frac{2}{\pi}}\,E\Big(\sup_{t\in T}\Big|\frac{1}{n_0-1}\sum_{j=1}^{n_0-1}\varepsilon_j\cdot f(\xi_j,t)\Big|\Big)$$

for all $n\ge n_0\ge 1$. Now, given $\varepsilon>0$, take $n_0\ge 1$ such that the last two expressions are less than ε . Then take the limit superior on both sides as $n\to\infty$. This clearly establishes (3.25) and finishes the proof of the equivalence (3.24)-(3.25). The proof of Theorem 3.8 is complete. □

3.4 The VC law of large numbers

The main purpose of this section is to prove the theorem of Vapnik and Chervonenkis which appeared in [89] and [90]. The theorem characterizes the uniform law of large numbers in terms of the asymptotic behaviour of the random entropy numbers. The sufficiency part of the proof is based on Rademacher randomization (presented in Section 3.3) and subgaussian inequality (3.1). We learned it from [98]. The necessity part of the proof relies upon Gaussian randomization (presented in Section 3.3) and Sudakov's minoration (3.19) (via an argument due to Talagrand which is based on Stirling's formula as recorded in [22]). The elements of this part are found in [32] and [22]. The results of this section will be extended and generalized in Sections 3.7–3.9 below.

1. Throughout we consider a sequence of independent and identically distributed random variables $\xi=\{\,\xi_j\mid j\ge 1\,\}$ defined on the probability space $(\Omega,\mathcal{F},P)$ with values in the measurable space $(S,\mathcal{A})$ and a common distribution law π . We suppose that a set T is given, as well as a map $f:S\times T\to\mathbf{R}$ such that $s\mapsto f(s,t)$ is π-measurable for all $t\in T$. We moreover assume that $|f(s,t)|\le 1$ for all $(s,t)\in S\times T$, and put $M(t)=\int_S f(s,t)\;\pi(ds)$ for $t\in T$. (This assumption will later be extended to more general cases in Remark 3.13 below.) Through the whole section we suppose that the structure of f and T is flexible enough, such that we have as much measurability as needed. (We recall that this approach can be supported in quite a general setting by using theory of analytic spaces as explained in Paragraph 5 of Introduction.)

2. In order to perform the Rademacher randomization we suppose that a Rademacher sequence $\varepsilon=\{\varepsilon_j\mid j\ge 1\}$ is given (see Section 3.1), which is defined on the probability space $(\Omega_\varepsilon,\mathcal{F}_\varepsilon,P_\varepsilon)$. Whenever the sequences ξ and ε appear together, we assume that they are both defined on the product probability space $(\Omega\otimes\Omega_\varepsilon,\mathcal{F}\otimes\mathcal{F}_\varepsilon,P\otimes P_\varepsilon)$, and thus may be understood mutually independent. Similarly, in order to perform the Gaussian randomization we suppose that a standard Gaussian sequence $g=\{g_j\mid j\ge 1\}$ is given, meaning that g_j for $j\ge 1$ are independent standard Gaussian variables (see Section 3.2), which is defined on the probability space $(\Omega_g,\mathcal{F}_g,P_g)$. Also, whenever the sequences ξ and g appear together, we assume that they are both defined on the product probability space $(\Omega\otimes\Omega_g,\mathcal{F}\otimes\mathcal{F}_g,P\otimes P_g)$, and thus may be understood mutually independent. Finally, whenever the three sequences ξ , ε and g appear together, we assume that they are all defined on the product probability space $(\Omega\otimes\Omega_\varepsilon\otimes\Omega_g,\mathcal{F}\otimes\mathcal{F}_\varepsilon\otimes\mathcal{F}_g,P\otimes P_\varepsilon\otimes P_g)$, and thus may be understood mutually independent.

3. The key novel concept in this section is the concept of a random entropy number. Setting $\mathcal{F} = \{ f(\,\cdot\,,t) \mid t \in T \}$ we define the *random entropy number* $N_n(\varepsilon, \mathcal{F})$ associated with ξ through $\mathcal{F}$ as the smallest number of open balls of radius $\varepsilon > 0$ in the sup-metric of $\mathbf{R}^n$ needed to cover the set $F_n = \{ \big(f(\xi_1,t),\dots,f(\xi_n,t)\big) \mid t \in T \}$ with $n \geq 1$. Since F_n is a random set in the cube $[-1,1]^n$, then $N_n(\varepsilon,\mathcal{F})$ is a random variable, bounded above by the constant $([1/\varepsilon]+1)^n$ for $n \geq 1$ and $\varepsilon > 0$ (recall the Definition in the beginning of Chapter 2). We are now ready to state the main result of this section.

Theorem 3.11 (Vapnik and Chervonenkis 1981)

Let $\xi = \{ \xi_j \mid j \geq 1 \}$ be a sequence of independent and identically distributed random variables. Then the uniform law of large numbers is valid:

$$\sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^n f(\xi_j,t) - M(t) \right| \longrightarrow 0 \quad P\text{-a.s.} \tag{3.32}$$

as $n \to \infty$, if and only if the condition is satisfied:

$$\lim_{n \to \infty} \frac{E\big(\log N_n(\varepsilon,\mathcal{F})\big)}{n} = 0 \tag{3.33}$$

for all $\varepsilon > 0$.

Proof. Sufficiency: Suppose that (3.33) is satisfied. Then by Corollary 2.8 and Theorem 3.8 it is enough to show that:

$$\sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot f(\xi_j,t) \right| \longrightarrow 0 \quad \text{in } P \otimes P_\varepsilon\text{-probability} \tag{3.34}$$

as $n \to \infty$. For this, put $N = N_n(\varepsilon,\mathcal{F})$ for $n \geq 1$ and $\varepsilon > 0$ given and fixed. Then by definition of N there exist vectors $Z_1 = (Z_{1,1},\dots,Z_{1,n}),\ \dots\ , Z_N = (Z_{N,1},\dots,Z_{N,n})$ in $[-1,1]^n$ and a map $i : T \to \{1,\dots,N\}$ such that:

$$\sup_{t \in T} \max_{1 \leq j \leq n} \big| f(\xi_j,t) - Z_{i(t),j} \big| < \varepsilon \ .$$

Hence by the subgaussian inequality (3.1) we obtain:

$$\begin{aligned}
&P_\varepsilon \Big\{ \sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot f(\xi_j,t) \right| > 2\varepsilon \Big\} \\
&\leq P_\varepsilon \Big\{ \sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \Big(f(\xi_j,t) - Z_{i(t),j} \Big) \right| > \varepsilon \Big\} + P_\varepsilon \Big\{ \sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot Z_{i(t),j} \right| > \varepsilon \Big\} \\
&\leq N \max_{1 \leq i \leq N} P_\varepsilon \Big\{ \left| \sum_{j=1}^n \varepsilon_j \cdot Z_{i,j} \right| > n\varepsilon \Big\} \leq 2N \max_{1 \leq i \leq N} \exp\left(\frac{-n^2\varepsilon^2}{2\sum_{j=1}^n |Z_{i,j}|^2} \right) \\
&\leq 2N \exp\left(\frac{-n\,\varepsilon^2}{2} \right) .
\end{aligned} \tag{3.35}$$

Now, put $A = \big\{ \sup_{t \in T} \big| \frac{1}{n} \sum_{j=1}^n \varepsilon_j f(\xi_j,t) \big| > 2\varepsilon \big\}$, then by (3.35) and Markov's inequality we get:

$$P\otimes P_\varepsilon\Big\{ \sup_{t\in T} \Big| \frac{1}{n}\sum_{j=1}^n \varepsilon_j \cdot f(\xi_j,t) \Big| > 2\varepsilon \Big\} = (P\otimes P_\varepsilon)(A) = \int_\Omega P_\varepsilon(A_\omega)\, P(d\omega)$$
$$= \int_{\{ \log N \le n\varepsilon^2/4 \}} P_\varepsilon(A_\omega)\, P(d\omega) \,+ \int_{\{ \log N > n\varepsilon^2/4 \}} P_\varepsilon(A_\omega)\, P(d\omega)$$
$$\le\, 2\exp\Big(\frac{-n\,\varepsilon^2}{4}\Big) + P\big\{ \log N > n\,\varepsilon^2/4 \big\} \le 2\exp\Big(\frac{-n\,\varepsilon^2}{4}\Big) + \frac{4}{\varepsilon^2}\,\frac{1}{n} E\big(\log N\big)\ .$$

Letting $n \to \infty$ we obtain (3.34) from (3.33), and the proof of sufficiency is complete.

Necessity: Consider the process:

$$X_{t,n} \,=\, \frac{1}{\sqrt{n}}\, \sum_{j=1}^n g_j \cdot f(\xi_j,t)$$

in $t\in T$ for $n\ge 1$ given and fixed. Then $X_n = \big(X_{t,n}\big)_{t\in T}$ is a Gaussian process on $(\Omega_g,\mathcal{F}_g,P_g)$ with the intrinsic pseudo-metric given by:

$$d_{X_n}(t',t'') \,=\, \frac{1}{\sqrt{n}}\Big(\sum_{j=1}^n \big|f(\xi_j,t')-f(\xi_j,t'')\big|^2 \Big)^{1/2}$$

for $t',t''\in T$. By Sudakov's minoration estimate (3.19) in Theorem 3.5 we have:

$$\varepsilon\,\sqrt{\log N\big(T,d_{X_n};\varepsilon\big)} \le C\,\frac{1}{\sqrt{n}}\, E_g\Big(\sup_{t\in T} \Big| \sum_{j=1}^n g_j\cdot f(\xi_j,t) \Big| \Big)$$

for $\varepsilon>0$ given and fixed. Integrating both sides with respect to P , we get:

$$\varepsilon\, E\Big(\sqrt{\frac{1}{n}\log N\big(T,d_{X_n};\varepsilon\big)} \Big) \,\le\, C\, E\Big(\sup_{t\in T} \Big| \frac{1}{n}\sum_{j=1}^n g_j\cdot f(\xi_j,t) \Big| \Big)\ .$$

Hence, if (3.32) is satisfied, then by Corollary 2.8 and Theorem 3.8 we get:

$$E\Big(\sqrt{\frac{1}{n}\log N\big(T,d_{X_n};\varepsilon\big)}\ \Big) \,\longrightarrow\, 0 \tag{3.36}$$

as $n\to\infty$. Since $d_{X_n}(t',t'') \le \max_{1\le j\le n} |f(\xi_j,t')-f(\xi_j,t'')|$, then we have $N(T,d_{X_n};\varepsilon) \le N_n(\varepsilon,\mathcal{F}) \le ([1/\varepsilon]+1)^n$, and thus $\frac{1}{n}\log N(T,d_{X_n};\varepsilon) \le \log([1/\varepsilon]+1)$ for all $n\ge 1$. This fact shows that the sequence of random variables $\{ \frac{1}{n}\log N(T,d_{X_n};\varepsilon) \mid n\ge 1 \}$ is uniformly integrable, and therefore (3.36) implies:

$$\frac{1}{n} E\Big(\log N\big(T,d_{X_n};\varepsilon\big)\Big) \,\longrightarrow\, 0 \tag{3.37}$$

as $n\to\infty$. The proof will be completed as soon as we show that (3.37) implies (3.33).

For this, let $0<\alpha<\varepsilon<1/2$ and $n\ge 1$ be given and fixed. By definition of $N = N(T,d_{X_n};\alpha\varepsilon/2)$, there is a map $\theta : T\to T$ such that $d_{X_n}(t,\theta(t)) < \alpha\varepsilon/2$ for all $t\in T$, and such that $card\,\{ \theta(t) \mid t\in T \} = N$. We recall that $\mathcal{F} = \{ f(\,\cdot\,,t) \mid t\in T \}$, and let $\mathcal{G} = \{ f(\,\cdot\,,t) - f(\,\cdot\,,\theta(t)) \mid t\in T \}$. Since for $t',t''\in T$ with $\theta(t')=\theta(t'')$ we have $f(\,\cdot\,,t')-f(\,\cdot\,,t'') = \big(f(\,\cdot\,,t')-f(\,\cdot\,,\theta(t'))\big) - \big(f(\,\cdot\,,t'')-f(\,\cdot\,,\theta(t''))\big)$ which is a difference of two elements from $\mathcal{G}$, it is clear that we have:

$$N_n(\varepsilon, \mathcal{F}) \le N(T, d_{X_n}; \varepsilon) \cdot N_n(\varepsilon, \mathcal{G}) \ . \tag{3.38}$$

In addition, we want to estimate $N_n(\varepsilon, \mathcal{G})$. For this, note if $g(\cdot, t) = f(\cdot, t) - f(\cdot, \theta(t))$ is from $\mathcal{G}$ for some $t \in T$, then by Jensen's inequality we get $\frac{1}{n}\sum_{j=1}^n |g(\xi_j, t)| \le (\frac{1}{n}\sum_{j=1}^n |g(\xi_j, t)|^2)^{1/2} = d_{X_n}(t, \theta(t)) < \alpha\varepsilon/2$. Thus there are at most $m = [\alpha n]$ members j from $\{1, \dots, n\}$ for which $|g(\xi_j, t)| \ge \varepsilon/2$. This fact motives us to define $\mathcal{H}$ as the family of all functions from $\{\xi_1, \dots, \xi_n\}$ into $\mathbf{R}$ which are zero at $n-m$ points ξ_k for which $|g(\xi_k, t)| < \varepsilon/2$, and which take values $k\varepsilon/2$ at the remaining m points for $k \in \mathbf{Z}$ with $|k| \le 4/\varepsilon$. Then we clearly get:

$$\min_{h \in \mathcal{H}} \max_{1 \le j \le n} \left| g(\xi_j, t) - h(\xi_j) \right| \le \varepsilon/2 \ .$$

It shows that $N_n(\varepsilon, \mathcal{G}) \le card\,(\mathcal{H})$, from which by counting the number of elements in $\mathcal{H}$ we find:

$$N_n(\varepsilon, \mathcal{G}) \le \binom{n}{m} \left(1 + \frac{8}{\varepsilon}\right)^m \ . \tag{3.39}$$

Now, by Stirling's formula ($n! = \theta_n \sqrt{2\pi}\, n^{n+1/2} e^{-n}$ with $\exp(1/12n+1) < \theta_n < \exp(1/12n)$ for all $n \ge 1$) one can easily verify the following fact:

$$\limsup_{n\to\infty} \frac{1}{n} \log\binom{n}{m} \le \alpha\,|\log\alpha\,| + (1-\alpha)\,|\log(1-\alpha)\,| \tag{3.40}$$

for all $0 < \alpha < 1$, where $m = [\alpha n]$. Combining (3.37)-(3.40) we conclude:

$$\limsup_{n\to\infty} \frac{1}{n} E\Big(N_n(\varepsilon, \mathcal{F})\Big) \le \alpha\,|\log\alpha\,| + (1-\alpha)\,|\log(1-\alpha)\,| + \alpha\,\log\left(1 + \frac{8}{\varepsilon}\right)$$

which is valid for all $0 < \alpha < \varepsilon < 1/2$. Letting $\alpha \downarrow 0$, we obtain (3.33) and complete the proof.

□

Remark 3.12

In the notation of Theorem 3.11 we could observe from the proof (see (3.37)) that the following two statements are equivalent (to the uniform law of large numbers (3.32)):

$$\lim_{n\to\infty} \frac{E\big(\log N_n(\varepsilon, \mathcal{F})\big)}{n} = 0 \tag{3.41}$$

$$\lim_{n\to\infty} \frac{E\big(\log N(T, d_{X_n}; \varepsilon\big)}{n} = 0 \tag{3.42}$$

for all $\varepsilon > 0$. Simple comparison arguments based on Jensen's inequality and the uniform boundedness of $\mathcal{F}$ shows that the pseudo-metric d_{X_n} (which appears naturally in the proof as the intrinsic pseudo-metric of the Gaussian process obtained by the procedure of randomization) may be equivalently replaced in (3.42) by any of the pseudo-metrics:

$$d_{X_n,p}(t', t'') = \left(\frac{1}{n}\sum_{j=1}^n \left| f(\xi_j, t') - f(\xi_j, t'')\right|^p\right)^{1\wedge 1/p}$$

for $t', t'' \in T$ with $0 < p < \infty$ and $n \ge 1$. It is the approach taken in [32] and [22]. Other pseudo-metrics with similar purposes appear in the literature as well.

Remark 3.13

It is assumed in Theorem 3.11 that $|f(s,t)| \leq 1$ for all $(s,t) \in S \times T$. The result extends to the case where the integrability condition is satisfied:

$$\sup_{t \in T} |f(\,\cdot\,,t)| \in L^1(\pi)\ . \tag{3.43}$$

To see this, let us define the map $F_T(s) = \sup_{t \in T} |f(s,t)|$ for $s \in S$, and put:

$$f_R(s,t) \,=\, f(s,t) \cdot 1_{\{\,F_T \leq R\}}(s)$$

for $(s,t) \in S \times T$ and $R > 0$. Then by Markov's inequality we find:

$$\begin{aligned} &P \otimes P_\varepsilon \Big\{ \sup_{t \in T} \Big| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot f(\xi_j, t) \Big| > \varepsilon \Big\} \\ &\leq P \otimes P_\varepsilon \Big\{ \sup_{t \in T} \Big| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot f(\xi_j, t) \cdot 1_{\{\,F_T \leq R\}}(\xi_j) \Big| > \frac{\varepsilon}{2} \Big\} \\ &\quad + P \otimes P_\varepsilon \Big\{ \sup_{t \in T} \Big| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot f(\xi_j, t) \cdot 1_{\{\,F_T > R\}}(\xi_j) \Big| > \frac{\varepsilon}{2} \Big\} \\ &\leq P \otimes P_\varepsilon \Big\{ \sup_{t \in T} \Big| \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cdot f_R(\xi_j, t) \Big| > \frac{\varepsilon}{2} \Big\} + \frac{2}{\varepsilon} E\Big(|F_T| \cdot 1_{\{\,F_T > R\}} \Big) \end{aligned}$$

for all $n \geq 1$, all $\varepsilon > 0$, and all $R > 0$. Since by (3.43) the last term tends to zero as $R \to \infty$, we see by Corollary 2.8 and Theorem 3.8 that $\mathcal{F} = \{\, f(\,\cdot\,,t) \mid t \in T\,\}$ satisfies the uniform law of large numbers (3.32), if and only if $\mathcal{F}_R = \{\, f_R(\,\cdot\,,t) \mid t \in T\,\}$ does so for all $R > 0$.

Thus, if we define the entropy numbers $N_{n,R}(\varepsilon, \mathcal{F}) := N_n(\varepsilon, \mathcal{F}_R)$ as in Paragraph 3 above (note that the unit cube $[-1,1]^n$ in the definition may be replaced by the cube $[-R,R]^n$ without any problem), then we see that under (3.43) the uniform law of large numbers (3.32) holds if and only if the condition is satisfied:

$$\lim_{n \to \infty} \frac{E\big(\log N_{n,R}(\varepsilon, \mathcal{F})\big)}{n} \,=\, 0 \tag{3.44}$$

for all $\varepsilon > 0$ and all $R > 0$. This completes the claim of the remark.

The VC Theorem 3.11 has also been extended to more general U-statistics of i.i.d. sequences in [2]. We emphasize that U-statistics are the most important example of reversed martingales. They are also treated by a different method in the Supplement.

3.5 The VC classes of sets

The purpose of this section is to present the concept of VC class of sets, which provide the best known (and investigated) examples satisfying the VC theorem (Theorem 3.11).

Let S be a set, and let $\mathcal{C} \subset 2^S$ be a family of subsets of S . Let us put:

$$\Delta_{\mathcal{C}}(A) \,=\, card\,\big(A \cap \mathcal{C}\big)$$

whenever $A \subset S$ is finite, where we set $A \cap \mathcal{C} = \{ A \cap C \mid C \in \mathcal{C} \}$. We say that $\mathcal{C}$ *shatters* A, if $\Delta_{\mathcal{C}}(A) = 2^{\,card\,(A)}$. Let us define:

$$m_{\mathcal{C}}(n) = \max \{ \Delta_{\mathcal{C}}(A) \mid A \subset S \,,\, card\,(A) = n \}$$

for $n \geq 1$. Finally, let us put:

$$V(\mathcal{C}) = \inf \{ n \geq 1 \mid m_{\mathcal{C}}(n) < 2^n \}$$

with $\inf(\emptyset) = +\infty$. The family $\mathcal{C}$ is called a *VC class* of sets, if $V(\mathcal{C}) < +\infty$.

Note that if $V(\mathcal{C}) < +\infty$, then $m_{\mathcal{C}}(n) < 2^n$ for all $n \geq V(\mathcal{C})$. Thus, the map $m_{\mathcal{C}}(n)$ is not of the (strict) exponential growth 2^n as $n \to \infty$. But then, as it is shown in the next theorem, it must be of a polynomial growth. This is the most striking fact about the VC classes of sets. The proof below is taken and adapted from [54].

Theorem 3.14
If $\mathcal{C}$ is a VC class of sets, then we have:

$$m_{\mathcal{C}}(n) \leq n^{V(\mathcal{C})} \tag{3.45}$$

for all $n \geq V(\mathcal{C})$.

Proof. Below we will prove a more general fact which implies:

$$card\,\big(A \cap \mathcal{C}\big) \leq card\,\{ B \subset A \mid card\,(B) < V(\mathcal{C}) \} \tag{3.46}$$

whenever $A \subset S$ is finite. Applying (3.46) to $A \subset S$ with $card(A) = n \geq V(\mathcal{C})$, we easily find:

$$\Delta_{\mathcal{C}}(A) \leq \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{V(\mathcal{C})-1} \leq 1 + n^{V(\mathcal{C})-1} \,. \tag{3.47}$$

This clearly establishes (3.45). In order to prove (3.46) we will show:

$$card\,\big(\mathcal{D}\big) \leq card\,\{ B \subset A \mid \mathcal{D} \text{ shatters } B \} \tag{3.48}$$

which is valid for any family $\mathcal{D} \subset 2^A$. It should be noted that (3.46) follows from (3.48) by taking $\mathcal{D} = A \cap \mathcal{C}$.

To prove (3.48) we define a map $T_x : \mathcal{D} \to 2^A$ for $x \in A$ as follows. We put $T_x(D) = D \backslash \{x\}$ if $D \setminus \{x\} \notin \mathcal{D}$, and we put $T_x(D) = D$ otherwise. First, we observe that T_x is one-to-one on $\mathcal{D}$. For this, let $T_x(D_1) = T_x(D_2)$. If $T_x(D_1) \in \mathcal{D}$, then $T_x(D_1) = D_1 = T_x(D_2) \in \mathcal{D}$ and thus $T_x(D_2) = D_2$. If $T_x(D_1) \notin \mathcal{D}$, then $T_x(D_1) = D_1 \setminus \{x\} = T_x(D_2) \notin \mathcal{D}$ and $x \in D_1$. Thus $T_x(D_2) = D_2 \setminus \{x\}$ and $x \in D_2$. In both cases it follows $D_1 = D_2$, and the first claim is proved. Consequently, we have:

$$card\,\big(T_x(\mathcal{D})\big) = card\,(\mathcal{D}) \tag{3.49}$$

for any $x \in A$. Second, we verify the fact:

(3.50) If $T_x(\mathcal{D})$ shatters given $B \subset A$, then $\mathcal{D}$ shatters B as well.

For this, let $C \subset B$ be given. Then there exists $D \in \mathcal{D}$ such that $C = B \cap T_x(D)$. If $x \notin B$, then $B \cap T_x(D) = B \cap D$. If $x \in C$, then $T_x(D) = D$. Finally, if $x \in B \setminus C$, then there exists $E \in \mathcal{D}$ such that $C \cup \{x\} = B \cap T_x(E)$. Thus $T_x(E) = E$ and therefore $E \setminus \{x\} \in \mathcal{D}$. Moreover $C = B \cap (E \setminus \{x\})$, and (3.50) follows. Third, we define $w(\mathcal{E}) = \sum_{E \in \mathcal{E}} card\,(E)$ whenever $\mathcal{E} \subset 2^A$, and note that $T_x(\mathcal{D}) \neq \mathcal{D}$ implies $w(T_x(\mathcal{D})) < w(\mathcal{D})$. Thus applying inductively T_x to $\mathcal{D}$ with $x \in A$ we obtain $\mathcal{D}_0 \subset 2^A$ with $w(\mathcal{D}_0)$ being minimal. In other words if $D \in \mathcal{D}_0$, then $D \setminus \{x\} \in \mathcal{D}_0$ for all $x \in A$. Thus if $D \in \mathcal{D}_0$ and $C \subset D$, then $C \in \mathcal{D}_0$ as well. Hence $\mathcal{D}_0$ shatters each of its own elements, and thus (3.48) is obvious for $\mathcal{D}_0$. Moreover by (3.49) we have $card\,(\mathcal{D}) = card\,(\mathcal{D}_0)$. Finally, by (3.50) we see that $\mathcal{D}$ shatters more sets than $\mathcal{D}_0$. Thus (3.48) follows for $\mathcal{D}$ as well, and the proof is complete. □

Remark 3.15

It is easily observed that in the notation of the VC theorem (Theorem 3.11) we have:

$$N_n(\varepsilon, \mathcal{F}_\mathcal{C}) \leq m_\mathcal{C}(n)$$

for all $n \geq 1$ and all $\varepsilon > 0$, where $\mathcal{C}$ is any family of subsets of the set S and where we denote $\mathcal{F}_\mathcal{C} = \{ 1_C \mid C \in \mathcal{C} \}$. Combining this fact with (3.45), we see that the entropy numbers $N_n(\varepsilon, \mathcal{F}_\mathcal{C})$ associated with a VC class $\mathcal{C}$ are of a *polynomial growth* when $n \to \infty$ for every $\varepsilon > 0$.

Corollary 3.16 (Extended GC Theorem 1.10)

Let $\xi = \{ \xi_j \mid j \geq 1 \}$ be a sequence of independent and identically distributed random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ with values in the measurable space $(S, \mathcal{A})$. If $\mathcal{C} \subset 2^S$ is a VC class of sets, then the uniform law of large numbers is valid:

$$\sup_{C \in \mathcal{C}} \left| \frac{1}{n} \sum_{j=1}^{n} 1_{\{\xi_j \in C\}} - P\{ \xi_1 \in C \} \right| \longrightarrow 0 \quad P\text{-a.s.}$$

as $n \to \infty$.

Proof. By Theorem 3.14 and Remark 3.15 we have:

$$0 \leq \frac{1}{n} E\Big(\log N_n(\varepsilon, \mathcal{F}_\mathcal{C}) \Big) \leq V(\mathcal{C}) \, \frac{\log n}{n} \to 0$$

as $n \to \infty$, where $\mathcal{F}_\mathcal{C} = \{ 1_C \mid C \in \mathcal{C} \}$. Thus the claim follows from Theorem 3.11. □

The examples of VC classes will be given in Section 3.10.1 below.

3.6 The Eberlein lemma

In order to generalize and extend the VC law of large numbers to stationary random variables in the next section, we will use a lemma due to Eberlein in [23], which may be interpreted as a decoupling inequality with an error term. The present section is devoted to its statement and proof.

Lemma 3.17 (Eberlein 1984)

Let $X = \{ X_j \mid j \geq 1\}$ be a sequence of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ with values in the measurable space $(S, \mathcal{A})$, and let $Y = \{ Y_j \mid j \geq 1\}$ be a sequence of random variables defined on a probability space $(\Omega', \mathcal{F}', P')$ with values in $(S, \mathcal{A})$ which is *coupled to the sequence* X *by the relation:*

$$\mathcal{L}\big(Y_1,\dots,Y_{\sigma_n w_n}\big) = \mathcal{L}\big(X_1,\dots,X_{w_n}\big) \otimes \mathcal{L}\big(X_{2w_n+1},\dots X_{3w_n}\big)$$
$$\otimes \dots \otimes \mathcal{L}\big(X_{(2\sigma_n-2)w_n+1},\dots,X_{(2\sigma_n-1)w_n}\big)$$

for σ_n , $w_n \geq 1$ *with* $n \geq 1$. *Let us put* $\sigma_1^l = \sigma(X_1,\dots,X_l)$ *and* $\sigma_l^\infty = \sigma(X_{l+1},X_{l+2},\dots)$ *for* $l \geq 1$, *and let us define the numbers:*

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

for $k \geq 1$. *Then the following estimate is valid:*

$$\Big| P\big\{ (X_1,\dots,X_{w_n},X_{2w_n+1},\dots,X_{3w_n},\ \dots\ ,X_{(2\sigma_n-2)w_n+1},\dots,X_{(2\sigma_n-1)w_n}) \in B \big\}$$
$$- P'\big\{ (Y_1,\dots,Y_{w_n},Y_{2w_n+1},\dots,Y_{3w_n},\ \dots\ ,Y_{(2\sigma_n-2)w_n+1},\dots,Y_{(2\sigma_n-1)w_n}) \in B \big\} \Big|$$
$$\leq (\sigma_n - 1)\,\beta_{w_n} \tag{3.51}$$

for all measurable sets $B \in \mathcal{A}^{\sigma_n w_n}$ *with* $n \geq 1$.

Proof. The proof is carried out by induction in $i = 1,\dots,\sigma_n$ where $n \geq 1$ is given and fixed. For $\sigma_n = 2$, we denote $U_1 = (X_1,\dots,X_{w_n})$, $U_2 = (X_{2w_n+1},\dots,X_{3w_n})$, $V_1 = (Y_1,\dots,Y_{w_n})$ and $V_2 = (Y_{2w_n+1},\dots,Y_{3w_n})$. Then we claim that:

$$\big|\, P\{\, (U_1,U_2) \in A \,\} - P'\{\, (V_1,V_2) \in A \,\} \,\big| \leq \beta_{w_n} \tag{3.52}$$

for all $A \in \mathcal{A}^{2w_n}$.

To prove (3.52), let $\mathcal{G}$ denote the family of all sets $A \in A^{2w_n}$ satisfying (3.52), and let $\mathcal{H}$ be the family of all sets from $\mathcal{A}^{2w_n}$ of the form $\cup_{i=1}^N B_i^1 \times B_i^2$ with $B_i^1 \times B_i^2 \in \mathcal{A}^{w_n} \otimes \mathcal{A}^{w_n}$ being disjoint for $i = 1,\dots,N$ and $N \geq 1$. Then $\mathcal{G}$ is a monotone class of sets, and $\mathcal{H}$ is an algebra of sets. Thus by the monotone class lemma (see 19]) we have $\sigma(\mathcal{G}) = \mathcal{M}(\mathcal{H})$. Hence, it is enough to show $\mathcal{H} \subset \mathcal{G}$. For this, notice that any element A from $\mathcal{H}$ allows the representation $A = \cup_{i=1}^N \cup_{j=1}^{M_i} B_i^1 \times B_{k_j}^2$ where $B_i^1 \in \mathcal{A}^{w_n}$ are disjoint for $i = 1,\dots,N$, and where $B_{k_j}^2 \in \mathcal{A}^{w_n}$ are disjoint for $j = 1,\dots,M$ with $M_i \leq M$ for all $i = 1,\dots,N$. Hence, by independence of V_1 and V_2 , the facts $V_1 \sim U_1$, $V_2 \sim U_2$, $\{U_1 \in B_i^1\} \in \sigma_1^{w_n}$ and $\{U_2 \in B_{k_j}^2\} \in \sigma_{2w_n}^\infty$, and the definition of β_{w_n} , we get:

$$\big|\, P\{\, (U_1,U_2) \in A \,\} - P'\{\, (V_1,V_2) \in A \,\} \,\big|$$
$$\leq \sum_{i=1}^N \sum_{j=1}^{M_i} \big|\, P\{\, U_1 \in B_i^1 \,,\, U_2 \in B_{k_j}^2 \} - P\{\, U_1 \in B_i^1 \,\} \cdot P\{\, U_2 \in B_{k_j}^2 \} \,\big| \leq \beta_{w_n} \,.$$

Thus, the proof of (3.52) is complete.

In addition, for $\sigma_n = 3$, consider $U_3 = (X_{4w_n+1},\dots,X_{5w_n})$ and $V_3 = (Y_{4w_n+1},\dots,Y_{5w_n})$. Then we claim that:

$$\text{(3.53)} \qquad \Big| \, P\{ \, (U_1, U_2, U_3) \in A \, \} - P'\{ \, (V_1, V_2, V_3) \in A \, \} \, \Big| \leq 2\beta_{w_n}$$

for all $A \in \mathcal{A}^{3w_n}$.

To prove (3.53), again, by the monotone class lemma it is enough to show (3.53) for $A = \cup_{i=1}^{N} \cup_{j=1}^{M_i} B_i^1 \times B_{k_j}^2$ where $B_i^1 \in \mathcal{A}^{w_n}$ are disjoint for $i = 1, \ldots, N$, and where $B_{k_j}^2 \in \mathcal{A}^{2w_n}$ are disjoint for $j = 1, \ldots, M$ with $M_i \leq M$ for all $i = 1, \ldots, N$. For this, first note that (3.52) holds as well (the arguments in the proof remain unchanged) if (U_1, U_2) is replaced by (U_2, U_3) , and (V_1, V_2) is replaced by (V_2, V_3) . By this fact, independence and equal distribution, and the definition of β_{w_n} , we get:

$$\begin{aligned}
&\Big| \, P\{ \, (U_1, U_2, U_3) \in A \, \} - P'\{ \, (V_1, V_2, V_3) \in A \, \} \, \Big| \\
&= \Bigg| \sum_{i=1}^{N} \sum_{j=1}^{M_i} P\{ \, U_1 \in B_i^1 \, , \, (U_2, U_3) \in B_{k_j}^2 \} - P\{ \, U_1 \in B_i^1 \, \} \cdot P\{ \, (U_2, U_3) \in B_{k_j}^2 \} \\
&\qquad + P\{ \, U_1 \in B_i^1 \, \} \cdot P\{ \, (U_2, U_3) \in B_{k_j}^2 \} - P\{ \, U_1 \in B_i^1 \, \} \cdot P'\{ \, (V_2, V_3) \in B_{k_j}^2 \} \Bigg| \\
&\leq \sum_{i=1}^{N} \sum_{j=1}^{M_i} \Big| P\{ \, U_1 \in B_i^1 \, , \, (U_2, U_3) \in B_{k_j}^2 \} - P\{ \, U_1 \in B_i^1 \, \} \cdot P\{ \, (U_2, U_3) \in B_{k_j}^2 \} \Big| + \\
&\qquad + \sum_{i=1}^{N} P\{ U_1 \in B_i^1 \} \cdot \Big| P\big\{ (U_2, U_3) \in \bigcup\nolimits_{j=1}^{M_i} B_{k_j}^2 \big\} - P'\big\{ (V_2, V_3) \in \bigcup\nolimits_{j=1}^{M_i} B_{k_j}^2 \big\} \Big| \leq 2\beta_{w_n} \, .
\end{aligned}$$

Thus, the proof of (3.53) is complete.

From (3.52) and (3.53) we see that (3.51) is valid for σ_n being 2 and 3. It is moreover clear that the procedure of passing from (3.52) to (3.53) could be inductively repeated as many times as one wishes, and in this way we can reach any given number σ_n . This fact completes the proof.

□

Remark 3.18

Note that our proof above shows that the sequence $X = \{ \, X_j \mid j \geq 1 \, \}$ in Lemma 3.17 may be arbitrary, thus not necessarily from the identical distribution.

3.7 The uniform ergodic theorem for absolutely regular dynamical systems

The aim of this section is to generalize and extend the VC law of large numbers (Theorem 3.11) to dynamical systems (see Paragraph 4 in Section 1.1). It turns out that this extension is possible for absolutely regular dynamical systems (defined below). The method of proof relies upon a blocking technique (which goes back to Bernstein [5]), a decoupling inequality due to Eberlein (Lemma 3.17), Rademacher randomization (Section 3.3) and the subgaussian inequality for Rademacher averages (Theorem 3.1). The result is generalized and extended to semi-flows and (non-linear) operators in the next section. All of this material is taken from [68].

1. Throughout, let $\{ X_i \}_{i \geq 1}$ be a *stationary* sequence of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$, with values in the measurable space $(S, \mathcal{A})$ and a common distribution law π , and with distribution law μ in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$. More precisely, this means that:

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

for all $1 \leq n_1 < \ldots < n_k$ and all $p \geq 1$ (see Paragraph 1 in Section 2.2). 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:

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

is ergodic with respect to μ (see Paragraph 2 in Section 2.2).

2. For every $l \geq 1$ introduce the σ-algebras:

$$\sigma_1^l = \sigma(X_1, \dots, X_l) \ \ \& \ \ \sigma_l^\infty = \sigma(X_{l+1}, X_{l+2}, \dots) \ .$$

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

$$\beta_k = \beta_k\big(\{ X_i \}_{i \geq 1}\big) = \sup_{l \geq 1} \int \sup_{A \in \sigma_{k+l}^\infty} |P(A \,|\, \sigma_1^l) - P(A)| \, dP \tag{3.54}$$

for all $k \geq 1$. Equivalently, the β-mixing coefficients may be defined as follows (see [10]):

$$\begin{aligned} \beta_k = \sup \Big\{ & \textstyle\sum_{i=1}^I \sum_{j=1}^J |P(A_i \cap B_j) - P(A_i) \cdot P(B_j)| : \\ & (A_i)_{i=1}^I \text{ is any finite partition in } \sigma_1^l \text{, and} \\ & (B_j)_{j=1}^J \text{ is any finite partition in } \sigma_{k+l}^\infty \text{ for } I, J, l \geq 1 \Big\} \end{aligned} \tag{3.55}$$

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 [92]-[93] who attribute it to Kolmogorov. Conditions for the absolute regularity of Gaussian stationary processes can be found in [76] p.180-190.

3. 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 Section 2.2). Thus, by Birkhoff's Theorem 1.6, or equivalently (2.12) above, if $X_1 \in L^1(P)$ then the following strong law of large numbers (SLLN) 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 (see (2.9) above) 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 $i \geq 1$.

Our wish is to extend this SLLN and obtain a uniform SLLN over a class $\mathcal{F}$ of real valued functions on S . Although the uniform SLLN in the general setting of stationarity is characterized in Section 2.2, our goal here is to provide an analog of the classic VC law of large numbers (Theorem 3.11) in the setting of absolute regularity. This approach involves conditions on the entropy number for $\mathcal{F}$. Recall that (see Paragraph 3 in Section 3.4) 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 $\big(f(X_1), \dots, f(X_n)\big)$ where f ranges over $\mathcal{F}$, and where $n \geq 1$ is given and fixed.

4. The main result of this section may now be stated as follows. In Section 3.8 below we will see (by deriving a uniform ergodic theorem for non-linear operators) that a P-probability version of the next theorem holds for sequences of random variables which are not stationary. Throughout we denote $a_n = o(b_n)$ to indicate that $a_n/b_n \to 0$ as $n \to \infty$.

Theorem 3.19

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

$$\frac{\beta_{w_n}}{w_n} = o(n^{-1}) \tag{3.56}$$

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

$$\lim_{n \to \infty} w_n \frac{E \log N_n^X(\varepsilon, \mathcal{F})}{n} = 0 \tag{3.57}$$

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

$$\sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} \big(f(X_i) - Ef(X_i) \big) \Big| \longrightarrow 0 \ \ P\text{-}a.s. \tag{3.58}$$

as $n \to \infty$.

Remark 3.20

It is easily verified that condition (3.56) with some sequence $w_n = o(n)$ is equivalent to the following condition:

$$\tfrac{n}{2} - w_n < \sigma_n w_n \leq \tfrac{n}{2} \quad \& \quad \sigma_n \beta_{w_n} \to 0 \tag{3.59}$$

where $\sigma_n = [n/2w_n]$ for $n \geq 1$. It turns out that (3.59) is precisely a version of condition (3.56) which will be used in the proof below. When X_i's 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 (3.56) is satisfied. Moreover, the weighted entropy condition (3.57) in this case reduces to the classical VC entropy condition. In this way we recover the sufficiency part of the VC theorem (Theorem 3.11). Finally, since the sequence $\{ \beta_k \}_{k \geq 1}$ is decreasing, it is no restriction to assume in Theorem 3.19 that $w_n \to \infty$, for otherwise the β-mixing coefficients β_k 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 (3.56) holds. Then we have:

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

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

Remark 3.21

When $\mathcal{F}$ consists of the indicators of sets from a VC class, by Remark 3.15 we see that the conclusion (3.58) of Theorem 3.19 holds whenever there exists a sequence $w_n = o(n)$, such that:

$$n \frac{\beta_{w_n}}{w_n} \to 0 \quad \& \quad w_n \frac{\log n}{n} \to 0 \tag{3.60}$$

as $n \to \infty$. For example, consider the case when *the mixing rate* r_β is strictly positive, where we recall that $r_\beta = \sup \{ r \geq 0 \mid \{n^r \beta_n\}_{n\geq 1}$ 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:

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

as $n \to \infty$. Thus, when $\mathcal{F}$ consists of the indicators of sets from a VC class, then the uniform SLLN holds whenever the mixing rate is strictly positive.

5. Before proving Theorem 3.19 we shall establish some preliminary facts. The proof is centered around the *blocking technique* which is described as follows. Setting $\sigma_n = [n/2w_n]$ for $n \geq 1$, divide the sequence $(X_1, \ldots, X_n)$ into $2\sigma_n$ blocks of length w_n , leaving a remainder block of length $n - 2\sigma_n w_n$. Define blocks:

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

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

$$R = \{ \; i \;\mid\; 2\sigma_n w_n + 1 \leq i \leq n \; \}$$

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

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

for all $n \geq 1$. Then the Eberlein lemma (Lemma 3.17) 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. It easily implies that for any bounded measurable function $g : S^{\sigma_n w_n} \to \mathbf{R}$ we have the decoupling estimate:

$$\begin{aligned} &\Big| Eg\,(X_1, \ldots, X_{w_n}, X_{2w_n+1}, \ldots, X_{3w_n}, \; \ldots \; , X_{(2\sigma_n-2)\, w_n+1}, \ldots, X_{(2\sigma_n-1) w_n}) \\ &\qquad - Eg\,(Y_1, \ldots, Y_{w_n}, Y_{2w_n+1}, \ldots, Y_{3w_n}, \; \ldots \; , Y_{(2\sigma_n-2)\, w_n+1}, \ldots, Y_{(2\sigma_n-1) w_n}) \Big| \\ &\leq (\sigma_n - 1) \cdot \beta_{w_n} \cdot \| \, g \, \|_\infty \end{aligned} \tag{3.61}$$

for all $n \geq 1$.

6. We now provide a proof of Theorem 3.19. As shown in the next section, the method of proof is flexible and admits a generalization to the non-stationary setting.

Proof of Theorem 3.19: By Corollary 2.8, it is enough to show convergence in P-probability in (3.58). 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 the Eberlein lemma (Lemma 3.17) to show that the entropy hypothesis (3.57) implies an entropy result for $\mathcal{F}$ with respect to the coupled block sequence $\{ Y_i \}_{i\geq 1}$. For this, the following definition is needed.

Definition. Let $\hat{N}_{\sigma_n}^X(\varepsilon,\mathcal{F})$ 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 $\mathbf{R}^{\sigma_n}$ with coordinates $f(X_i)$ for $i = 1, 2w_n+1, 4w_n+1 \ldots, (2\sigma_n-2)w_n+1$ formed by $f \in \mathcal{F}$. Define $\hat{N}_{\sigma_n}^Y(\varepsilon,\mathcal{F})$ in a similar way by replacing X_i with Y_i.

We now show that the entropy condition:

$$\lim_{n\to\infty} \Big(\frac{1}{\sigma_n} E \log \hat{N}_{\sigma_n}^X(\varepsilon,\mathcal{F}) \Big) = 0 \tag{3.62}$$

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

$$\lim_{n\to\infty} \Big(\frac{1}{\sigma_n} E \log \hat{N}_{\sigma_n}^Y(\varepsilon,\mathcal{F}) \Big) = 0 \tag{3.63}$$

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:

$$\frac{1}{\sigma_n} \log \hat{N}_{\sigma_n}^Z(\varepsilon,\mathcal{F}) \leq \frac{1}{\sigma_n} \log \big(C/\varepsilon\big)^{\sigma_n} = \log\big(C/\varepsilon\big)$$

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

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

where Z equals X or Y, respectively. Moreover $\|g\|_\infty \leq \log\big(C/\varepsilon\big)$, and thus by (3.59) and (3.61) we obtain:

$$\Big| \frac{1}{\sigma_n} E \log \hat{N}_{\sigma_n}^X(\varepsilon,\mathcal{F}) - \frac{1}{\sigma_n} E \log \hat{N}_{\sigma_n}^Y(\varepsilon,\mathcal{F}) \Big| \leq (\sigma_n - 1)\, \beta_{w_n} \log\big(C/\varepsilon\big) \to 0$$

as $n \to \infty$. This shows the desired equivalence of (3.62) and (3.63).

Moreover, we note that (3.57) trivially implies (3.62), and therefore *the entropy condition (3.57) implies the entropy condition (3.63).* We will use this heavily in the next step.

Step 2. In this step we use the Eberlein lemma (Lemma 3.17) and condition (3.63) 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:

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

for all $\varepsilon > 0$, and all $n \geq 1$. For the last equality above we use (3.59) from which we obtain:

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

as $n \to \infty$. Hence by stationarity and decoupling (Lemma 3.17), we obtain:

$$\begin{aligned} (3.64) \qquad & P\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} f(X_i) \Big| > \varepsilon \Big\} \\ & \leq 2\, P\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{\sigma_n} \sum_{i \in B_j} f(X_i) \Big| > \varepsilon/4 \Big\} + o(1) \\ & \leq 2\, Q\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{\sigma_n} \sum_{i \in B_j} f(Y_i) \Big| > \varepsilon/4 \Big\} + o(1) \end{aligned}$$

for all $\varepsilon > 0$ and all $n \geq 1$, since $(\sigma_n - 1)\, \beta_{w_n} = o(1)$ by (3.59).

To conclude, it suffices to show that the last term in (3.64) becomes small as n increases. Since the random variables $\sum_{i \in B_j} f(Y_i)$ are independent (and identically distributed) for $1 \leq j \leq \sigma_n$, it is enough to show that the symmetrized version of the last term in (3.64) becomes arbitrarily small when n increases. Thus by standard symmetrization lemmas (Corollary 2.8 and Theorem 3.8) it is enough to show that for $\varepsilon > 0$ given and fixed, there exists $n_\varepsilon \geq 1$ such that:

$$(3.65) \qquad (Q \otimes Q_\varepsilon)\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{\sigma_n} \varepsilon_j \cdot \sum_{i \in B_j} f(Y_i) \Big| > \varepsilon \Big\} \leq \varepsilon$$

for all $n \geq n_\varepsilon$, where $\{ \varepsilon_j \}_{j \geq 1}$ is a Rademacher sequence defined on a probability space $(\Lambda_\varepsilon, \mathcal{G}_\varepsilon, Q_\varepsilon)$ 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:

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

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:

$$(3.66) \qquad (Q \otimes Q_\varepsilon)\Big\{ \sup_{f \in \mathcal{F}} \Big| \frac{1}{\sigma_n} \sum_{j=1}^{\sigma_n} \varepsilon_j f(Y_{(2j-2)w_n+1}) \Big| > 2\varepsilon \Big\} < 2\varepsilon$$

for all $n \geq n_\varepsilon$.

To show (3.66), proceed as in the necessity part of the proof of Theorem 3.11. Assume without loss of generality that $\mathcal{F}$ has the uniform bound 1 . Let A denote the event:

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

with $\varepsilon > 0$ and $n \geq 1$ fixed. Observe that (3.63) implies the existence of $n_\varepsilon \geq 1$ such that:

$$E\log \hat{N}^{Y}_{\sigma_n}(\varepsilon,\mathcal{F}) \,\leq\, \varepsilon^4\sigma_n$$

$$\exp\big(\sigma_n\varepsilon^2(\varepsilon-1/2)\big) \,\leq\, \varepsilon/2$$

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

By the definition of the entropy number $N = \hat{N}^{Y}_{\sigma_n}(\varepsilon,\mathcal{F})$, there are vectors x_l in $[-1\,,1\,]^n$ for $1 \leq l \leq N$ with coordinates $x_{l,i}$ for $i = 1, 2w_n+1, 4w_n+1, \ldots, (2\sigma_n-2)w_n+1$, 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 $1, 2w_n+1, 4w_n+1, \ldots, (2\sigma_n-2)w_n+1$. By the triangle inequality we have:

$$(3.67)\qquad (Q\otimes Q_\varepsilon)(A) \leq (Q\otimes Q_\varepsilon)\Big\{ \sup_{f\in\mathcal{F}} \frac{1}{\sigma_n}\Big| \sum_{j=1}^{\sigma_n} \varepsilon_j\, f(Y_{(2j-2)w_n+1}) - \sum_{j=1}^{\sigma_n}\varepsilon_j\, x_{l(f),(2j-2)w_n+1}\Big| > \varepsilon \Big\} + \; (Q\otimes Q_\varepsilon)\Big\{ \max_{1\leq l\leq N} \frac{1}{\sigma_n}\Big| \sum_{j=1}^{\sigma_n}\varepsilon_j\, x_{l,(2j-2)w_n+1}\Big| > \varepsilon \Big\}$$

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 (3.67) is zero, by choice of $x_{l(f)}$. Applying the subgaussian inequality (Theorem 3.1) to the second term yields:

$$(3.68)\qquad (Q\otimes Q_\varepsilon)\Big\{ \max_{1\leq l\leq N} \frac{1}{\sigma_n}\Big| \sum_{j=1}^{\sigma_n}\varepsilon_j\, x_{l,(2j-2)w_n+1}\Big| > \varepsilon \Big\} \leq 2N\cdot \exp\Big(\frac{-\sigma_n^2\,\varepsilon^2}{2\,\sigma_n}\Big) \,=\, 2N\cdot \exp\Big(\frac{-\sigma_n\,\varepsilon^2}{2}\Big)\ .$$

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

$$(3.69)\qquad Q\Big\{\, \log N \,\geq\, \sigma_n\varepsilon^3 \,\Big\} \,\leq\, \varepsilon\ .$$

Finally, combining (3.67)–(3.69), the left hand side of (3.66) becomes:

$$(Q\otimes Q_\varepsilon)(A) = \int_A 1_{\{N\geq \exp(\sigma_n\varepsilon^3)\}}d(Q\otimes Q_\varepsilon) + \int_A 1_{\{N< \exp(\sigma_n\varepsilon^3)\}}d(Q\otimes Q_\varepsilon)$$
$$\leq \varepsilon + 2\cdot\exp\Big(\sigma_n\varepsilon^3\Big)\cdot\exp\Big(\frac{-\sigma_n\varepsilon^2}{2}\Big) \,\leq\, 2\varepsilon$$

for all $n \geq n_\varepsilon$. This completes Step 2 and the proof of Theorem 3.19. $\square$

7. In the remainder of this section we first extend Theorem 3.19 to the unbounded case. Since this approach follows in a straightforward way along the lines of Remark 3.13, we will not provide all details. After this, we formulate Theorem 3.19 in terms of dynamical systems.

It is assumed in Theorem 3.19 that the elements $f \in \mathcal{F}$ satisfy $\|f\|_\infty \leq C$, for some $C > 0$. 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:

$$f_R(s) \,=\, f(s) \cdot 1_{\{F_{\mathcal{F}} \leq R\}}(s)$$

for $s \in S$. 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 $n \geq 1$ is given and fixed. With this notation, we may now state a generalization of Theorem 3.19 as follows.

Theorem 3.22
Let $\{X_i\}_{i \geq 1}$ be an absolutely regular sequence of random variables satisfying the condition:

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

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 (3.58).

Proof. Follow the proof of Theorem 3.19. In Step 2 observe that by Chebyshev's inequality:

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

for all $R > 0$ and all $n \geq 1$. Letting $n \to \infty$, Theorem 3.19 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. $\square$

8. Finally, we formulate the result of Theorem 3.19 in terms of dynamical systems. Throughout, let $(\Omega, \mathcal{F}, \mu, T)$ be a given *dynamical system*. Thus $(\Omega, \mathcal{F}, \mu)$ is a probability space, and T is a measure-preserving transformation of Ω (see Paragraph 4 in Section 1.1). Let $\kappa : \Omega \to S$ be a measurable function, where $(S, \mathcal{A})$ is a measurable space. For every $l \geq 1$ 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 by:

$$\beta_k = \beta_k(\kappa) = \sup_{l \geq 1} \int \sup_{A \in \sigma_{k+l}^\infty} |\, \mu(A \,|\, \sigma_1^l) - \mu(A) \,| \ d\mu \tag{3.70}$$

for all $k \geq 1$. The dynamical system $(\Omega, \mathcal{F}, \mu, T)$ 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 in the beginning of this section, 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. 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. By $N_n(\varepsilon, \mathcal{F}, \kappa)$ we 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 all vectors in $\mathbf{R}^n$ of the form $\big(f(\kappa), f(\kappa \circ T), \dots, f(\kappa \circ T^{n-1})\big)$ formed by $f \in \mathcal{F}$, where $n \geq 1$ is given and fixed. The next result shows that a weighted VC entropy condition insures that $\mathcal{F} \in UET(\kappa)$.

Theorem 3.23

Let $(\Omega, \mathcal{F}, \mu, T)$ *be an absolutely regular dynamical system through a factorization* $\kappa : \Omega \to S$ *satisfying the condition:*

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

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. The claim follows from Theorem 3.19 upon identifying the random variables X_i with $\kappa \circ T^{i-1}$ for $i \geq 1$. This completes the proof. $\square$

From Theorem 3.22 it clearly follows that Theorem 3.23 admits an extension to the case of unbounded $\mathcal{F}$, the details of which are left to the reader. Clearly, there are also other ways to extend and generalize Theorem 3.19 (or Theorem 3.23). 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. A review of similar results obtained for ϕ-mixing and α-mixing sequences is found in [97]. With a little more care one can also obtain rates of convergence in Theorem 3.19 (see [97] for such results).

3.8 Extension to semi-flows and (non-linear) operators

In the first part of this section we extend the basic Theorem 3.19 to one-parameter semi-flows (Theorem 3.24), and in the second part we extend it to (non-linear) operators (Theorem 3.25).

1. By a *one-parameter semi-flow* $\{ T_t \}_{t \geq 0}$ defined on the probability space $(\Omega, \mathcal{F}, \mu)$, 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. Note that (1.10) is then well-defined μ-a.s. by Fubini's theorem (see [19]). 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 $t \geq 0$ (see Section 1.1).

As above, let $(S, \mathcal{A})$ 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:

$$(3.71) \qquad \sup_{f \in \mathcal{F}} \Big| \frac{1}{\Delta} \int_0^\Delta f(\kappa \circ T_t) \; dt \; - \int_\Omega f(\kappa(\omega)) \; \mu(d\omega) \Big| \longrightarrow 0 \quad \mu\text{-a.s. } \omega \in \Omega$$

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

$$(3.72) \qquad \kappa(\omega') = \kappa(\omega'') \;\Rightarrow\; \int_0^1 f\big(\kappa \circ T_t(\omega')\big)\, dt = \int_0^1 f\big(\kappa \circ T_t(\omega'')\big)\, dt$$

whenever $\omega', \omega'' \in \Omega$ and $f \in \mathcal{F}$. Under assumption (3.72) 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 $f \in \mathcal{F}$, 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 . Setting $T := T_1$, the (β, κ)-mixing coefficient $\beta_k := \beta_k(\kappa)$ of T for $k \geq 1$ is defined as in (3.70). 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.

Theorem 3.24

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

$$\frac{\beta_{w_n}(\kappa)}{w_n} = o(n^{-1}) \tag{3.73}$$

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* $\mathcal{F}$ *satisfies the uniform ergodic theorem (3.71).*

Proof. Recalling the reduction principle (1.10)-(1.11), the claim follows from Theorem 3.23 together with the following two facts:

$$\frac{1}{N}\int_0^{N-1} f\big(\kappa\circ T_t(\omega)\big)\, dt = \frac{1}{N}\sum_{i=0}^{N-1} F(\kappa\circ T^i(\omega), f) \quad \text{for all } \omega\in\Omega \tag{3.74}$$

$$\sup_{f\in\mathcal{F}} \Big| \frac{1}{N}\int_0^{N-1} f\big(\kappa\circ T_t(\omega)\big)\, dt - \frac{1}{\Delta}\int_0^{\Delta} f\big(\kappa\circ T_t(\omega)\big)\, dt \Big| \longrightarrow 0 \quad \text{for all } \omega\in\Omega \tag{3.75}$$

as $N := [\Delta] \to \infty$. This completes the proof. □

It is clear that Theorem 3.24 admits an extension to the case of unbounded $\mathcal{F}$ having an envelope belonging to $L^1(\mu)$ (see Theorem 3.22). We will not pursue this, but instead consider an extension of Theorem 3.19 to (non-linear) operators. In the process we will see that a convergence in probability version of Theorem 3.19 actually holds for sequences of random variables which are neither identically distributed nor stationary.

2. Throughout, let $(\Omega, \mathcal{F}, \mu)$ 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:

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=0}^{n-1} \Big(f\big(T^i(g)(\omega)\big) - \int_\Omega f\big(T^i(g)(\omega)\big)\ \mu(d\omega)\Big) \Big| = 0 \tag{3.76}$$

in μ-probability, as $n\to\infty$. This result may be interpreted as a pointwise uniform ergodic theorem for the operator T (see Section 2.5.5).

We note that if the operator T is induced by means of a measure-preserving transformation, then (3.76) reduces to the setting considered in Theorem 3.23 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 3.19 and Theorem 3.23.

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 $(\Omega, \mathcal{F}, \mu)$ 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 (3.76) holds. In this case we write $\mathcal{F} \in UET(g)$.

It turns out that the methods employed in Section 3.7, 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 . Recall that $N_n(\varepsilon, \mathcal{F}, g)$ 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 3.25

Suppose that the measure space $(\Omega, \mathcal{F}, \mu)$ 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 T with respect to g satisfy:

$$\beta_{w_n}(g) = o(n^{-1}) \tag{3.77}$$

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

$$\lim_{n \to \infty} \ w_n \ \frac{E \log N_n(\varepsilon, \mathcal{F}, g)}{n} \ = \ 0 \tag{3.78}$$

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

Remark 3.26

As the next proof shows, T can be a non-linear operator as well, that is, any map from $L^1(\mu)$ into $L^1(\mu)$.

Proof of Theorem 3.25. As already noted, the random variables:

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

do not form a stationary sequence, so Theorem 3.25 is not immediate from Theorem 3.19. Additionally, it does not seem possible to apply Corollary 2.8 to deduce the μ-a.s. convergence. Nonetheless, we may prove convergence to zero in μ-probability by adapting the methods used to prove Theorem 3.19. This is done as follows.

First, by Remark 3.18 we know that the Eberlein lemma (Lemma 3.17) holds for arbitrary sequences of random variables, thus for those which are not identically distributed as well. Therefore, letting $\{\, Y_i \,\}_{i \geq 1}$ be a sequence of random variables defined on a probability space $(\Lambda, \mathcal{G}, \nu)$ with independent blocks satisfying:

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

with $\sigma_n = [n/2w_n]$ for $n \geq 1$, we may modify the proof of Step 1 of Theorem 3.19 as follows.

Definition. Let $\hat{N}^X_{\sigma_n w_n}(\varepsilon, \mathcal{F})$ 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 $\mathbf{R}^{\sigma_n w_n}$ with coordinates $f(X_i)$ for $i = 1, \ldots, w_n, 2w_n{+}1, \ldots, 3w_n, \ \ldots \ , (2\sigma_n{-}2)w_n{+}1, \ldots, (2\sigma_n{-}1)w_n$ formed by $f \in \mathcal{F}$. Define $\hat{N}^Y_{\sigma_n w_n}(\varepsilon, \mathcal{F})$ in a similar way by replacing X_i with Y_i .

We now show that the entropy condition:

$$\lim_{n\to\infty} \left(\frac{w_n}{n} E \log \hat{N}^X_{\sigma_n w_n}(\varepsilon, \mathcal{F}) \right) = 0 \tag{3.79}$$

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

$$\lim_{n\to\infty} \left(\frac{w_n}{n} E \log \hat{N}^Y_{\sigma_n w_n}(\varepsilon, \mathcal{F}) \right) = 0 \tag{3.80}$$

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:

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

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

$$\begin{aligned} &Eg\,(Z_1, \ldots, Z_{w_n}, Z_{2w_n+1}, \ldots, Z_{3w_n}, \ \ldots \ , Z_{(2\sigma_n-2)\,w_n+1}, \ldots, Z_{(2\sigma_n-1)w_n}) \\ &\qquad = \frac{w_n}{n} E \log \hat{N}^Z_{\sigma_n w_n}(\varepsilon, \mathcal{F}) \end{aligned}$$

where Z equals X or Y , respectively. Moreover $\|g\|_\infty \leq w_n \log \big(C/\varepsilon\big)^{1/2}$, and thus by (3.61) and (3.77) we obtain:

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

as $n \to \infty$. This shows the desired equivalence of (3.79) and (3.80).

Moreover, we note that (3.78) trivially implies (3.79), and therefore *the entropy condition (3.78) implies the entropy condition (3.80).* We will use this heavily in the next step.

Concerning Step 2 of the proof of Theorem 3.19, we need to make the following modifications to the decoupling arguments:

$$\begin{aligned} &\mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{2\,\sigma_n w_n - 1} \Big(f\big(T^i(g)\big) - \int_\Omega f\big(T^i(g)(\omega)\big)\ \mu(d\omega)\Big) \Big| > \varepsilon/2 \Big\} \\ &\leq \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{\sigma_n} \sum_{i\in B_j} \Big(f\big(T^{i-1}(g)\big) - \int_\Omega f\big(T^{i-1}(g)(\omega)\big)\ \mu(d\omega)\Big) \Big| > \varepsilon/4 \Big\} \\ &\quad + \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{\sigma_n} \sum_{i\in \hat{B}_j} \Big(f\big(T^{i-1}(g)\big) - \int_\Omega f\big(T^{i-1}(g)(\omega)\big)\ \mu(d\omega)\Big) \Big| > \varepsilon/4 \Big\} \leq \end{aligned}$$

$$\text{(3.81)} \qquad \leq \nu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{j=1}^{\sigma_n}\sum_{i\in B_j} \big(f(Y_i) - Ef(Y_i)\big) \Big| > \varepsilon/4 \Big\}$$

$$+ \nu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{j=1}^{\sigma_n}\sum_{i\in \hat{B}_j} \big(f(Y_i) - Ef(Y_i)\big) \Big| > \varepsilon/4 \Big\} + 2(\sigma_n - 1)\,\beta_{w_n}$$

for all $n \geq 1$, where the last inequality follows by the Eberlein lemma (Lemma 3.17). Clearly, as B_j and $\hat{B}_j$ play symmetric roles, the first two terms in (3.81) 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 $1 \leq j \leq \sigma_n$, by standard symmetrization lemmas (explore the proof of (3.26) and observe that it does not use the identical distribution of the underlying sequence but only the independence) it is enough to show that for $\varepsilon > 0$ given and fixed, there exists $n_\varepsilon \geq 1$ such that:

$$\text{(3.82)} \qquad (\nu \otimes \nu_\varepsilon)\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{j=1}^{\sigma_n} \varepsilon_j \cdot \sum_{i\in B_j} f(Y_i) \Big| > 2\varepsilon \Big\} \leq 2\varepsilon$$

for all $n \geq n_\varepsilon$, where $\{\varepsilon_j\}_{j\geq 1}$ is a Rademacher sequence defined on a probability space $(\Lambda_\varepsilon, \mathcal{G}_\varepsilon, \nu_\varepsilon)$ 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 (3.82), proceed as in Step 2 of the proof of Theorem 3.19. Assume without loss of generality that $\mathcal{F}$ has the uniform bound 1. Let A denote the event:

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

with $\varepsilon > 0$ and $n \geq 1$ fixed. Observe that (3.80) implies the existence of $n_\varepsilon \geq 1$ such that:

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

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

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

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

$$\text{(3.83)} \qquad (\nu \otimes \nu_\varepsilon)(A) \leq (\nu \otimes \nu_\varepsilon)\Big\{ \sup_{f\in\mathcal{F}} \frac{1}{n}\Big| \sum_{j=1}^{\sigma_n} \varepsilon_j \cdot \sum_{i\in B_j} f(Y_i) - \sum_{j=1}^{\sigma_n} \varepsilon_j \cdot \sum_{i\in B_j} x_{l(f),i} \Big| > \varepsilon \Big\}$$

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

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

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

with the max as above. The first term on the right-hand side of inequality (3.83) is zero, by choice of $x_{l(f)}$. Applying the subgaussian inequality (Theorem 3.1) to the second term yields:

$$(3.84) \qquad (\nu \otimes \nu_\varepsilon) \Big\{ \max_{1 \le l \le N} \frac{1}{n} \Big| \sum_{j=1}^{\sigma_n} \varepsilon_j \cdot \sum_{i \in B_j} x_{l,i} \Big| > \varepsilon \Big\}$$
$$\le 2N \max_{1 \le l \le N} \exp\Big(\frac{-n^2 \varepsilon^2}{2} \Big(\sum_{j=1}^{\sigma_n} \Big(\sum_{i \in B_j} x_{l,i} \Big)^2 \Big)^{-1} \Big)$$
$$\le 2N \exp\Big(\frac{-n^2 \varepsilon^2}{2 \sigma_n w_n^2} \Big) \le 2N \exp\Big(\frac{-n\,\varepsilon^2}{w_n} \Big) .$$

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

$$(3.85) \qquad \nu \Big\{ \log N \ge \frac{n \varepsilon^3}{w_n} \Big\} \le \varepsilon .$$

Finally, combining (3.83)–(3.85), the left hand side of (3.82) becomes:

$$(\nu \otimes \nu_\varepsilon)(A) = \int_A 1_{\{N \ge \exp(n\varepsilon^3/w_n)\}} d(\nu \otimes \nu_\varepsilon) + \int_A 1_{\{N < \exp(n\varepsilon^3/w_n)\}} d(\nu \otimes \nu_\varepsilon)$$
$$\le \varepsilon + 2 \exp\Big(\frac{n\,\varepsilon^3}{w_n} \Big) \exp\Big(\frac{-n\,\varepsilon^2}{w_n} \Big) \le 2\varepsilon$$

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

It is clear that Theorem 3.25 admits an extension to the case of unbounded $\mathcal{F}$ having an envelope belonging to $L^1(\mu)$ (see Theorem 3.22). We will not pursue this in more detail, but instead pass to the problem of uniformity over factorizations in the next section.

3.9 Uniformity over factorizations

In this section we consider extensions of Theorem 3.23 and Theorem 3.24 to dynamical systems equipped with a family K of factorizations $\kappa : \Omega \to S$. We also consider a generalization of Theorem 3.25 which holds uniformly over a class $\mathcal{G}$ of functions from $L^1(\mu)$. Thus we adopt the notation from these theorems and present the results in Theorem 3.27, Theorem 3.28 and Theorem 3.29 below, respectively. (For applications of Theorem 3.27 and Theorem 3.29 see Example 3.37 and Example 3.38 below.) This material is taken from [68].

Theorem 3.27

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

$$(3.86) \qquad \sup_{\kappa \in \mathrm{K}} \Big(\frac{\beta_{w_n}(\kappa)}{w_n} \Big) = o(n^{-1})$$

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:*

$$\lim_{n\to\infty} \sup_{\kappa\in\mathrm{K}} \left(w_n \ \frac{E\log N_n(\varepsilon,\mathcal{F},\kappa)}{n} \right) = 0 \tag{3.87}$$

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

$$\sup_{\kappa\in\mathrm{K}} \ \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} f(\kappa\circ T^i) - \int_\Omega f(\kappa(\omega))\,\mu(d\omega) \Big| > \varepsilon \Big\} \longrightarrow 0 \tag{3.88}$$

$$\sup_{\kappa\in\mathrm{K}} \ E \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=0}^{n-1} f(\kappa\circ T^i) - \int_\Omega f(\kappa(\omega))\,\mu(d\omega) \Big| \longrightarrow 0 \tag{3.89}$$

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

Proof. The proof is essentially a modification of the proof of Theorem 3.19. First, given T , construct an associated coupled block sequence of random variables $\{ v_i \}_{i\ge 1}$ on a probability space $(\Lambda,\mathcal{G},\nu)$ with values in Ω and with the property:

$$\mathcal{L}(v_1,\ldots,v_{\sigma_n w_n}) = \bigotimes_1^{\sigma_n} \mathcal{L}(T^0,T^1,\ \ldots\ ,T^{w_n-1})$$

where $\sigma_n = [n/2w_n]$ for $n \ge 1$. Next, given $\kappa \in \mathrm{K}$, write:

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

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

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

for all $n \ge 1$. Following the argument in the proof of Theorem 3.19, we obtain the decoupled inequality as follows:

$$\begin{aligned} &\sup_{\kappa\in\mathrm{K}} \ \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^{n} \Big(f(X_i^\kappa) - \int_\Omega f(X_i^\kappa(\omega))\,\mu(d\omega)\Big) \Big| > \varepsilon \Big\} \\ &\le 2 \sup_{\kappa\in\mathrm{K}} \ \nu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n} \sum_{j=1}^{\sigma_n} \sum_{i\in B_j} \Big(f(Y_i^\kappa) - \int_\Lambda f(Y_i^\kappa(\lambda))\,\nu(d\lambda)\Big) \Big| > \varepsilon/4 \Big\} \\ &\quad + o(1) + \sup_{\kappa\in\mathrm{K}} (\sigma_n - 1)\beta_{w_n}(\kappa) \end{aligned} \tag{3.90}$$

for all $n \ge 1$. The last term in (3.90) is clearly $o(1)$ by hypothesis (3.86). The first term in (3.90) converges to zero as $n \to \infty$ by the methods of the proof of Theorem 3.19, together with the uniform entropy hypothesis (3.87), and the fact that the centering terms drop out when we randomize. This completes the proof of (3.88).

Finally, (3.89) 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. This completes the proof. □

Theorem 3.28

Let $\{T_t\}_{t\ge 0}$ be a measurable measure-preserving semi-flow of the probability space $(\Omega, \mathcal{F}, \mu)$, let $(S, \mathcal{A})$ be a measurable space, and let K be a family of measurable functions (factorizations) $\kappa : \Omega \to S$ satisfying the condition:

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

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

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

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 3.24 using the uniform approach of Theorem 3.27. We also make use of (3.74) and a uniformized version of (3.75) as follows:

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

as $N := [\Delta] \to \infty$. This completes the proof. □

Finally, we consider a generalization of Theorem 3.25 which holds uniformly over a class $\mathcal{G}$ of functions from $L^1(\mu)$. Again, as the following proof shows, T can be a non-linear operator, that is, an arbitrary map from $L^1(\mu)$ into $L^1(\mu)$. The result is stated as follows.

Theorem 3.29

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

$$\sup_{g \in \mathcal{G}} \beta_{w_n}(g) = o(n^{-1}) \tag{3.91}$$

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:

$$\lim_{n\to\infty} \sup_{g \in \mathcal{G}} \left(w_n \, \frac{E \log N_n(\varepsilon, \mathcal{F}, g)}{n} \right) = 0 \tag{3.92}$$

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

$$\text{(3.93)} \qquad \sup_{g\in\mathcal{G}} \mu\Big\{ \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=0}^{n-1} \Big(f(T^i(g)) - \int_\Omega f(T^i(g)(\omega))\, \mu(d\omega)\Big) \Big| > \varepsilon \Big\} \longrightarrow 0$$

$$\text{(3.94)} \qquad \sup_{g\in\mathcal{G}} E \sup_{f\in\mathcal{F}} \Big| \frac{1}{n}\sum_{i=0}^{n-1} \Big(f(T^i(g)) - \int_\Omega f(T^i(g)(\omega))\, \mu(d\omega)\Big) \Big| \longrightarrow 0$$

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

Proof. This follows along the lines of the proof of Theorem 3.27. The lack of stationarity may be overcome as in the proof of Theorem 3.25. This completes the proof. □

3.10 Examples and complements

This section contains examples and complementary facts which are aimed to support and clarify the results from the previous sections of this chapter.

3.10.1 Examples of VC classes of sets. In this section we present some typical examples of VC classes of sets. The first example stated below is classical and goes back to [34] and [13]. The second one is found in [28]. The third one is a consequence of the second one, and is initially due to Dudley (see [28]). The fourth one is from [73], and is due to Steele and Dudley. The last fact in this section (Proposition 3.34) helps us to build new VC classes from the existing ones.

We begin by recalling some notation from Section 3.5. Let S be a set, and let $\mathcal{C} \subset 2^S$ be a family of subsets of S . For $A \subset S$ finite, we denote $A \cap \mathcal{C} = \{ A \cap C \mid C \in \mathcal{C} \}$ and put $\Delta_{\mathcal{C}}(A) = card\,\big(A \cap \mathcal{C}\big)$. We say that $\mathcal{C}$ *shatters* A , if $\Delta_{\mathcal{C}}(A) = 2^{\,card\,(A)}$. We put $m_{\mathcal{C}}(n) = \max\{ \Delta_{\mathcal{C}}(A) \mid A \subset S \, , \, card\,(A) = n \}$ for $n \geq 1$, and denote $V(\mathcal{C}) = \inf\{ n \geq 1 \mid m_{\mathcal{C}}(n) < 2^n \}$ with $\inf(\emptyset) = +\infty$. The family $\mathcal{C}$ is called a *VC class* of sets, if $V(\mathcal{C}) < +\infty$ (see Corollary 3.16).

Example 3.30

Let $S = \mathbf{R}^n$ for $n \geq 1$, and let $\mathcal{C} = \{ 1_{]-\infty,x]} \mid x \in \mathbf{R}^n \}$. Then $\mathcal{C}$ is a VC class of sets with $V(\mathcal{C}) = 2$. It follows by a straightforward verification.

Example 3.31

Let $S = \mathbf{R}^n$ for $n \geq 1$, and let $\mathcal{D} \subset \mathcal{B}(\mathbf{R}^n)$ be a family of Borel sets satisfying the following *separation property*:

$$\text{(3.95)} \qquad co\,(D_1 \setminus D_2) \cap co\,(D_2 \setminus D_1) = \emptyset$$

for all $D_1 , D_2 \in \mathcal{D}$. Then $\mathcal{D}$ is a VC class of sets with $V(\mathcal{D}) \leq n+2$. (We clarify that $co\,(\cdot)$ denotes the convex hull of a set.)

Indeed, we shall show that $\mathcal{D}$ does not shatter any $A \subset \mathbf{R}^n$ with $card\,(A) = n+2$. Let A be such a set. Then by Radon's theorem (see [87]), there exist disjoint sets A_1 and A_2 in $\mathbf{R}^n$, such that $A = A_1 \cup A_2$ and $co\,(A_1) \cap co\,(A_2) \neq \emptyset$. So, if $\mathcal{D}$ does shatter A , then we can select D_1 and D_2 from $\mathcal{D}$, such that $A_1 = A \cap D_1$ and $A_2 = A \cap D_2$. Clearly $A_1 \subset D_1 \setminus D_2$ and $A_2 \subset D_2 \setminus D_1$. Hence we get:

$$co\,(D_1 \setminus D_2) \cap co\,(D_2 \setminus D_1) \supset co\,(A_1) \cap co\,(A_2) \neq \emptyset \ .$$

This contradicts hypothesis (3.95), and the proof of the claim is complete.

Example 3.32

Let $S = \mathbf{R}^n$ for $n \geq 1$, and let $\mathcal{B}$ be the family of all closed balls in $\mathbf{R}^n$. Then $\mathcal{B}$ is a VC class of sets with $V(\mathcal{B}) = n{+}2$. It follows from Example 3.31 together with the fact that there exists $A \subset \mathbf{R}^n$ with $card\,(A) = n{+}1$ which is shattered by $\mathcal{B}$.

Example 3.33

Let F be a finite dimensional vector space of real valued functions on the set S . Then $\mathcal{C} = \{\, \{f \geq 0\} \mid f \in F \,\}$ is a VC class of sets with $V(\mathcal{C}) \leq 1 + dim\,(F)$.

Put $d = dim\,(F)$, and let $S_0 = \{s_1, \ldots, s_{d+1}\} \subset S$ be the set consisting of $d{+}1$ arbitrary points from S . We show that $\mathcal{C}$ does not shatter S_0 . For this, define the linear operator $T : F \to \mathbf{R}^{d+1}$ by $T(f) = \big(f(s_1), f(s_2), \ldots, f(s_{d+1})\big)$ for $f \in F$. Then $dim\,\big(T(F)\big) \leq d$, and thus there exists a non-zero $\alpha = (\alpha_1, \ldots, \alpha_{d+1}) \in \mathbf{R}^{d+1}$ such that:

$$\big\langle T(f), \alpha \big\rangle \,=\, 0 \tag{3.96}$$

for all $f \in F$. Let $P = \{\, s_i \in S_0 \mid \alpha_i \geq 0 \,\}$ and $N = \{\, s_i \in S_0 \mid \alpha_i < 0 \,\}$. Replacing α by $-\alpha$ if needed, we may and do assume that $N \neq \emptyset$. Then from (3.96) we get:

$$\textstyle\sum_{s_i \in P} \alpha_i f(s_i) \,=\, \sum_{s_i \in N} (-\alpha_i) f(s_i) \tag{3.97}$$

where by definition the left-hand side reads zero if P is empty. Suppose now that $\mathcal{C}$ shatters S_0 . Then there exists $f \in F$ such that $P = S_0 \cap \{f \geq 0\}$. Inserting this fact into (3.97) we obtain a contradiction. This completes the proof of the claim.

Proposition 3.34

Let S and T be sets, and let $\mathcal{C}$, $\mathcal{D} \subset 2^S$ and $\mathcal{E} \subset 2^T$ be VC classes of sets. Then $\mathcal{C}^c = \{\, C^c \mid C \in \mathcal{C} \,\}$, $\mathcal{C} \cup \mathcal{D} = \{\, C \cup D \mid C \in \mathcal{C}\ , D \in \mathcal{D} \,\}$, $\mathcal{C} \cap \mathcal{D} = \{\, C \cap D \mid C \in \mathcal{C}\ , D \in \mathcal{D} \,\}$ and $\mathcal{C} \times \mathcal{E} = \{\, C \times E \mid C \in \mathcal{C}\ , E \in \mathcal{E} \,\}$ are VC classes of sets.

Proof. It follows straightforwardly by definition. □

3.10.2 Weber's counter-example.

The example is from [67] and shows that the inequality (3.26) may fail if the underlying sequence $\xi = \{\, \xi_j \mid j \geq 1 \,\}$ is only known to be strongly mixing (thus stationary and ergodic as well). It indicates that the Rademacher randomization, which appears as a straightforward tool in the proof of the VC theorem (Theorem 3.11), is not applicable to the strongly mixing case (thus not to the stationary and ergodic either). It is in accordance with our results from Section 3.7 which are valid for absolutely regular (β-mixing) sequences, thus being somewhere in between the i.i.d. case and the strongly mixing case. To describe more precisely this border of applicability appears worthy of consideration.

Example 3.35 (Weber 1993)

We show that the inequality (3.26) may fail if the underlying sequence $\{\, \xi_j \mid j \geq 1 \,\}$ is only assumed to be stationary and ergodic. For this we shall consider a simple case where T consists of a single point, and where f equals the identity map on the real line. The sequence $\{\, \xi_j \mid j \geq 1 \,\}$ itself is for a moment only assumed to be stationary and Gaussian. Thus our question reduces to verify the following inequality:

$$E\,\Big|\,\frac{1}{n}\sum_{j=1}^{n}\xi_j\,\Big| \le C\cdot E\,\Big|\,\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\cdot\xi_j\,\Big| \tag{3.98}$$

for all $n \ge 1$ with some constant $C > 0$, where $\{\,\varepsilon_j \mid j \ge 1\,\}$ is a Rademacher sequence independent of $\{\,\xi_j \mid j \ge 1\,\}$. Let us first consider the right-hand side of this inequality. For this denote by $\|\cdot\|_{\Psi_2}$ the Orlicz norm induced by the function $\Psi_2(x) = \exp(x^2) - 1$ for $x \in \mathbf{R}$. Then it is easily verified that we have $\|X\|_1 \le 6/5\,\|X\|_{\Psi_2}$ whenever X is a random variable. Hence by Kahane-Khintchine's inequality for $\|\cdot\|_{\Psi_2}$ (see [64]) and Jensen's inequality we get:

$$\begin{aligned} E\,\Big|\,\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\cdot\xi_j\,\Big| &= E_\xi\,E_\varepsilon\,\Big|\,\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\cdot\xi_j\,\Big| \le \frac{6}{5}\cdot\frac{1}{n}\cdot E_\xi\,\Big\|\sum_{j=1}^{n}\varepsilon_j\cdot\xi_j\Big\|_{\psi_2,\varepsilon} \\ &\le \frac{6}{5}\cdot\sqrt{\frac{8}{3}}\cdot\frac{1}{n}\cdot E_\xi\,\Big(\sum_{j=1}^{n}|\xi_j|^2\Big)^{1/2} \le 2\cdot\frac{1}{n}\sqrt{n\,E|\xi_1|^2} = \frac{2}{\sqrt{n}}\cdot\sqrt{E|\xi_1|^2} \end{aligned} \tag{3.99}$$

for all $n \ge 1$. On the other hand since $n^{-1}\sum_{j=1}^{n}\xi_j$ is Gaussian, then:

$$\Big(E\,\Big|\,\frac{1}{n}\sum_{j=1}^{n}\xi_j\,\Big|^2\Big)^{1/2} \le D\cdot E\,\Big|\,\frac{1}{n}\sum_{j=1}^{n}\xi_j\,\Big| \tag{3.100}$$

for some constant $D > 0$ and all $n \ge 1$. Inserting (3.99) and (3.100) into (3.98) we obtain:

$$E\,\Big|\sum_{j=1}^{n}\xi_j\Big|^2 \le G\cdot n\cdot E|\xi_1|^2 \tag{3.101}$$

for all $n \ge 1$ with $G = \sqrt{2CD}$. Thus it is enough to show that (3.101) may fail in general. Since $\{\,\xi_j \mid j \ge 1\,\}$ is stationary, then we have $E(\xi_i\xi_j) = R(i-j)$ for all $i, j \ge 1$. Moreover, it is easily verified that the left-hand side in (3.101) may be written as follows:

$$E\,\Big|\sum_{j=1}^{n}\xi_j\Big|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}E(\xi_i\xi_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}R(i-j) = n\,R(0) + \sum_{k=1}^{n-1}2\,(n-k)\,R(k) \tag{3.102}$$

for all $n \ge 1$. Let us in addition consider a particular case by putting $R(k) = 1/(k+1)$ for all $k \ge 0$. Then R is a convex function satisfying $R(0) = 1$ and $\lim_{n\to\infty} R(n) = 0$, and therefore by *Polya's theorem* (combine [11] p.241 with [56] p.70-71 & p.83) it is the covariance function of a centered stationary Gaussian sequence $\{\,\xi_j \mid j \ge 1\,\}$. Moreover, since $\lim_{n\to\infty} R(n) = 0$ then by *Maruyama's theorem* [57] (see [14] p.369) this sequence is strongly mixing, and thus ergodic as well. Finally, from (3.102) we easily obtain:

$$\begin{aligned} E\,\Big|\sum_{j=1}^{n}\xi_j\Big|^2 &= n + 2\sum_{k=1}^{n-1}\frac{n-k}{k+1} = 2\,(n+1)\sum_{k=1}^{n-1}\frac{1}{k+1} - n + 2 \\ &\ge 2\,(n+1)\int_1^n \frac{1}{x+1}\,dx - n + 2 \\ &= 2\,(n+1)\log(n+1) - (1 + \log 4)\,(n+1) + 3 \ge n\log n \end{aligned}$$

for all $n \ge 1$. This inequality contradicts (3.101), and therefore (3.98) is false in this case as well.

3.10.3 Talagrand's condition and Nobel's counter-example. In this section we first present Talagrand's necessary and sufficient condition for the uniform law of large numbers [84] (see also [85]), and then a counter-example due to Nobel [61] which shows that both the VC theorem (Theorem 3.11) and Talagrand's theorem may simultaneously fail in the stationary ergodic case.

1. Throughout we consider a sequence of independent and identically distributed random variables $\xi = \{ \xi_j \mid j \geq 1 \}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ with values in the measurable space $(S, \mathcal{A})$ and a common distribution law π . We suppose that a family $\mathcal{C} \subset 2^S$ is given, and we consider the uniform law of large numbers over $\mathcal{C}$ as follows:

$$\sup_{C \in \mathcal{C}} \left| \frac{1}{n} \sum_{j=1}^{n} 1_{\{\xi_j \in C\}} - P\{\xi_1 \in C\} \right| \longrightarrow 0 \quad P\text{-a.s.} \tag{3.103}$$

as $n \to \infty$. Recall that $\Delta_{\mathcal{C}}(A) = card\,(A \cap \mathcal{C})$ with $A \cap \mathcal{C} = \{ A \cap C \mid C \in \mathcal{C} \}$ whenever $A \subset S$ is finite, and $\mathcal{C}$ is said to shatter A , if $\Delta_{\mathcal{C}}(A) = 2^{\,card\,(A)}$. Moreover, if A is arbitrary (infinite) subset of S , then $\mathcal{C}$ is said to shatter A , if it shatters each finite subset of A .

Theorem A (Talagrand 1987) *The uniform law of large numbers (3.103) fails to hold, if and only if the condition is satisfied:*

(3.104) *There exists $A \in \mathcal{A}$ with $\pi(A) > 0$ for which the trace of π to A is non-atomic, such that $\mathcal{C}$ shatters the set $\{ \xi_{n_k} \mid k \geq 1\}$ consisting of those ξ_j's that lie in A , for P-almost every realization of the sequence $\xi = \{ \xi_j \mid j \geq 1 \}$.*

Another observation of interest in this section is that in the notation of Sections 3.4-3.5 we have the following identity:

$$\Delta_{\mathcal{C}}\big(\{\xi_1, \ldots, \xi_n\}\big) = N_n(\varepsilon, \mathcal{F}_{\mathcal{C}}) \tag{3.105}$$

for all $n \geq 1$ and all $0 < \varepsilon < 1/2$, where $\mathcal{F}_{\mathcal{C}} = \{ 1_C \mid C \in \mathcal{C} \}$. Thus, in the present case, the VC theorem (Theorem 3.11) may be reformulated as follows.

Theorem B (Vapnik and Chervonenkis 1971/81) *The uniform law of large numbers (3. 103) holds, if and only if the condition is satisfied:*

$$\lim_{n \to \infty} \frac{E\Big(\log \Delta_{\mathcal{C}}\big(\{\xi_1, \ldots, \xi_n\}\big)\Big)}{n} = 0 \ . \tag{3.106}$$

2. In the construction of Nobel's example below we make use of the Kakutani-Rokhlin lemma which is described as follows (see [36]). We consider a dynamical system $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is supposed to be a *Lebesgue space* (meaning that it is isomorphic (mod 0) to the ordinary Lebesgue space $([0,1], \mathcal{L}([0,1]), \lambda)$, which in turn means that there exist sets of measure zero $Z_1 \subset X$ and $Z_2 \subset [0,1]$, and a measurable bijection $\psi : X \setminus Z_1 \to [0,1] \setminus Z_2$, such that the inverse ψ^{-1} is measurable, and $\mu = \lambda \psi$), and T is supposed to be an *aperiodic* (measure-preserving) transformation (which means that $\mu\{ x \in X \mid T^n(x) = x \} = 0$ for all $n \geq 1$). It should be noted that every one-to-one ergodic (measure-preserving) transformation T of the Lebesgue space X is aperiodic. (This fact is, however, not true in general.)

(3.107) **(Kakutani-Rokhlin lemma)** *If T is an aperiodic (measure-preserving) transformation of the Lebesgue space $(X, \mathcal{B}, \mu)$, then for every $\varepsilon > 0$ and every $n \geq 1$ there exists* *a set* $A \in \mathcal{B}$ *such that the sets* A , $T^{-1}(A)$, ... , $T^{-(n-1)}(A)$ *are disjoint, and such that we have:*

$$\mu\big(A \cup T^{-1}(A) \cup \ldots \cup T^{-(n-1)}(A)\big) > 1-\varepsilon \ .$$

In this context, it is instructive to explore the dynamics of the trajectory $T^i(x)$ for $x \in T^{-j}(A)$ with $i \geq 1$ and $0 \leq j \leq n-1$.

Example 3.36 (Nobel 1992)

The example shows that there exists a stationary ergodic sequence $\xi = \{ \xi_j \mid j \geq 1 \}$ and a family of sets $\mathcal{C}$ for which the uniform law of large numbers (3.103) fails to hold, even though (3.104) fails as well, while (3.106) is fulfilled. The construction may be presented as follows.

Take any dynamical system $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a Lebesgue space (isomorphic (mod 0) to $[0,1]$), and where T is an one-to-one ergodic (measure-preserving) transformation of X . Then T is aperiodic, and thus the Kakutani-Rokhlin lemma (3.107) may be applied. In this way we obtain a sequence of sets $\{ A_p \mid p \geq 1 \}$ in X satisfying:

(3.108) $\quad A_p$, $T^{-1}(A_p)$, ... , $T^{-(p-1)}(A_p)$ are disjoint

(3.109) $\quad \mu\big(A_p \cup T^{-1}(A_p) \cup \ldots \cup T^{-(p-1)}(A_p)\big) > 1-1/p$

for all $p \geq 1$. We define $C_p = \bigcup_{i=0}^{[p/2]-1} T^{-i}(A_p)$ for all $p \geq 2$, and in this way we obtain the family $\mathcal{C} = \{ C_p \mid p \geq 2 \}$. The sequence $\xi = \{ \xi_j \mid j \geq 1 \}$ is defined in the usual way:

$$\xi_j(x) = T^{j-1}(x)$$

for $x \in X$ and $j \geq 1$. Then $\xi = \{ \xi_j \mid j \geq 1 \}$ is stationary and ergodic with values in X and the common law μ .

We first verify that (3.106) is fulfilled. For this, we establish the estimate:

(3.110) $\quad \Delta_{\mathcal{C}}\big(\{ x , T(x) , \ldots , T^{n-1}(x) \}\big) \leq 4n-2$

for all $x \in X$ and all $n \geq 1$. To do so, we recall (3.105) and consider the set:

$$F_n = \Big\{ \big(1_C(x) , 1_C(T(x)) , \ldots , 1_C(T^{n-1}(x))\big) \mid C \in \mathcal{C} \Big\}$$

with $n \geq 1$ given and fixed. Hence $\Delta_{\mathcal{C}}\big(\{ x , T(x) , \ldots , T^{n-1}(x) \}\big) = card\,(F_n)$. Therefore we explore $card\,(F_n)$. First, look at those C_p's for which $p \geq 2n$. Then the number of vectors:

$$\big(1_{C_p}(x) , 1_{C_p}(T(x)) , \ldots , 1_{C_p}(T^{n-1}(x))\big)$$

clearly does not exceed $2n$ when p runs through the set $\{ 2n , 2n+1 , \ldots \}$. Adding $2n-2$ remaining C_p's for $2 \leq p < 2n$, we obtain the estimate $card\,(F_n) \leq 2n + (2n-2)$. These facts complete the proof of (3.110). From (3.110) we obtain (3.106), and the first claim is complete.

Next, we show that (3.103) fails to hold. For this, fix $n \geq 1$ and consider the set A_{4n} . Then $\frac{1}{n} \sum_{j=0}^{n-1} 1_{C_{4n}}(T^j(x)) = 1$ for all $x \in B_n := \bigcup_{j=0}^{n-1} T^{-(j+n)}(A_{4n})$. Moreover, by construction we have $\mu(C_{4n}) \leq 1/2$. These two facts show that:

$$\sup_{C \in \mathcal{C}} \Big| \frac{1}{n} \sum_{j=0}^{n-1} 1_C(T^j(x)) - \mu(C) \Big| \geq \frac{1}{2}$$

for all $x \in B_n$. From (3.108)-(3.109) we find $4n\,\mu(A_{4n}) \geq 1-1/4n$, and thus we get:

$$\mu(B_n) \;=\; n\,\mu(A_{4n}) \geq \frac{1}{4}\Big(1-\frac{1}{4n}\Big) \geq \frac{3}{16}\;.$$

This shows that we do not have convergence in probability in (3.103), and the proof of the second claim is complete.

Finally, we show that (3.104) fails as well. For this, first note that in the notation of Theorem A above we have $(S,\mathcal{A},\pi) = (X,\mathcal{B},\mu)$. Now, suppose that there exists $A \in \mathcal{B}$ such that $\mathcal{C}$ shatters those points of the set $\{\, T^{j-1}(x) \mid j \geq 1 \,\}$ that lie in A . Let $B_n = \{\, x \in X \mid \frac{1}{n}\sum_{j=0}^{n-1} 1_A(T^j(x)) \geq \mu(A)/2 \,\}$ for $n \geq 1$. Then by Birkhoff's Ergodic Theorem 1.6 we have $\mu(B_n) \to 1$ as $n \to \infty$. Moreover, by the assumption we get:

$$\Delta_{\mathcal{C}}\big(\{\, x\,,\, T(x)\,,\,\dots\,,\, T^{n-1}(x)\,\}\big) \geq 2^{\sum_{j=0}^{n-1} 1_A(T^j(x))} \geq 2^{n\mu(A)/2}$$

for all $x \in B_n$ with $n \geq 1$. Hence we obtain:

$$\mu\Big\{\, x \in X \mid \Delta_{\mathcal{C}}\big(\{\, x\,,\, T(x)\,,\,\dots\,,\, T^{n-1}(x)\,\}\big) > 4n-2 \,\Big\} > 0$$

for $n \geq 1$ large enough. This contradicts (3.110) and completes the proof of the last claim.

3.10.4 Uniform convergence of moving averages. The section consists of two examples. In the first one (Example 3.37) we show how Theorem 3.27 applies to moving averages. In the second one (Example 3.38) we show how Theorem 3.29 applies to moving averages.

Example 3.37

Let $(\Omega,\mathcal{F},\mu,T)$ 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 $\kappa_m(s_1,s_2,\dots) = s_1+\dots+s_m$ 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 (3.86). Then it follows from (3.89) that we have:

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

as $n \to \infty$. For example, we may take $\mathcal{F}$ to be the family of indicators of sets in any VC class.

Example 3.38

Let $(\Omega,\mathcal{F},\mu)$ 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 (3.91). Then it follows from (3.94) 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 $n \to \infty$.

3.10.5 Metric entropy and majorizing measure type conditions. In Section 2.5.6 we have seen that metric entropy and majorizing measure type conditions (Theorems A and B in Section 2.5.6) could not be successfully applied for obtaining the uniform law of large numbers in the context of the Blum-DeHardt approach. In this section we indicate how these conditions *do* apply in the context of the VC approach (see [32] and [74]).

1. Throughout we consider a sequence of independent and identically distributed random variables $\xi = \{ \xi_j \mid j \geq 1 \}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ with values in the measurable space $(S, \mathcal{A})$ and a common distribution law π . We suppose that a set T and a map $f : S \times T \to \mathbf{R}$ are given, such that $s \mapsto f(s,t)$ is π-measurable for all $t \in T$, and put $\mathcal{F} = \{ f(\,\cdot\,,t) \mid t \in T \}$. We moreover assume that $|f(s,t)| \leq 1$ for all $(s,t) \in S \times T$, and put $M(t) = \int_S f(s,t)\, \pi(ds)$ for $t \in T$. The random entropy number $N_n(\varepsilon, \mathcal{F})$ associated with ξ through $\mathcal{F}$ is defined to be the smallest number of open balls of radius $\varepsilon > 0$ in the sup-metric of $\mathbf{R}^n$ needed to cover the random set $F_n = \{ \big(f(\xi_1,t), \ldots, f(\xi_n,t)\big) \mid t \in T \}$ with $n \geq 1$.

We are interested in obtaining the uniform law of large numbers as follows:

$$\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j, t) - M(t) \Big| \longrightarrow 0 \quad P\text{-a.s.} \tag{3.111}$$

as $n \to \infty$. According to Corollary 2.8 and Theorem 3.8, for (3.111) it is enough to show:

$$E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot f(\xi_j, t) \Big| \Big) \longrightarrow 0 \tag{3.112}$$

as $n \to \infty$, where $\varepsilon = \{ \varepsilon_j \mid j \geq 1 \}$ is a Rademacher sequence defined on the probability space $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon)$ and understood to be independent from $\xi = \{ \xi_j \mid j \geq 1 \}$.

2. In order to bound the left-hand side in (3.112), without loss of generality, we may and do assume that $0 \in \mathcal{F}$. (Otherwise, we could fix an element $f_0 \in \mathcal{F}$ and work with $\mathcal{F}_0 = \{ f - f_0 \mid f \in \mathcal{F} \}$, noting that $N_n(\varepsilon, \mathcal{F}) = N_n(\varepsilon, \mathcal{F}_0)$ for $n \geq 1$ and $\varepsilon > 0$.) Therefore we have:

$$E\Big(\sup_{t\in T} \Big| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot f(\xi_j, t) \Big| \Big) \leq E\Big(\sup_{t',t''\in T} \Big| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot \big(f(\xi_j, t') - f(\xi_j, t'') \big) \Big| \Big) \tag{3.113}$$

for all $n \geq 1$. In order to bound the right-side in (3.113), we shall use Theorem A from Section 2.5.6. For this, we define the process (in $t \in T$) on Ω_ε by:

$$X_t^n = \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot f(\xi_j(\omega), t)$$

for any given and fixed $\omega \in \Omega$ and $n \geq 1$. Let us take $\psi(x) = \exp(x^2) - 1$ for the Young function. Then by the Kahane-Khintchine inequality in the Orlicz space $L^{\psi}(P)$ (see [64]) we get:

$$\|X_s^n - X_t^n\|_{\psi} \leq \sqrt{\frac{8}{3}} \cdot \frac{1}{n} \cdot \left(\sum_{j=1}^{n} \big|f(\xi_j, s) - f(\xi_j, t)\big|^2 \right)^{1/2} \tag{3.114}$$

$$\leq \sqrt{\frac{8}{3}} \cdot \frac{1}{\sqrt{n}} \cdot \max_{1 \leq j \leq n} \big|f(\xi_j, s) - f(\xi_j, t)\big| := \sqrt{\frac{8}{3}} \cdot \frac{1}{\sqrt{n}} \cdot d_\infty^n(s,t)$$

for all $s, t \in T$ and all $n \geq 1$. Thus, the condition (2.171) is fulfilled with the process $X_t = \sqrt{3n/8}\, X_t^n(\omega)$ (for $t \in T$) and pseudo-metric $d = d_\infty^n(\omega)$, with $\omega \in \Omega$ and $n \geq 1$ given and fixed. Hence by (2.172), and the fact that $\log(1+x) \leq (\log 3 / \log 2) \log x$ for $x \geq 2$, we get:

$$E_\varepsilon \left(\sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \cdot f(\xi_j, t) \right| \right) \leq 8 \sqrt{\frac{8 \log 3}{3 \log 2}} \int_0^2 \sqrt{\frac{\log\big(N_n(\varepsilon, \mathcal{F})\big)}{n}}\, d\varepsilon$$

for all $n \geq 1$. Taking the P-integral and using (3.26), we obtain the inequality:

$$E\left(\sup_{t \in T} \left| \frac{1}{n} \sum_{j=1}^{n} f(\xi_j, t) - M(t) \right| \right) \leq C \int_0^2 E\left(\sqrt{\frac{\log\big(N_n(\varepsilon, \mathcal{F})\big)}{n}} \right) d\varepsilon \tag{3.115}$$

for all $n \geq 1$ with $C = 16\,(8 \log 3 / 3 \log 2)^{1/2}$. From (3.115) we see that (3.111) is valid as soon as we have:

$$\lim_{n \to \infty} \int_0^2 E\left(\sqrt{\frac{\log\big(N_n(\varepsilon, \mathcal{F})\big)}{n}} \right) d\varepsilon = 0 \,. \tag{3.116}$$

3. It is easily verified that the same method (where one uses Theorem B from Section 2.5.6 instead of Theorem A) shows that (3.111) is valid as soon as we have:

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} E\left(\inf_{m \in Pr(T, d_\infty^n)} \sup_{t \in T} \int_0^2 \sqrt{\log\left(1 + \frac{1}{m(B_{d_\infty^n}(t, \varepsilon))}\right)}\, d\varepsilon \right) = 0 \,. \tag{3.117}$$

Here $Pr(T, d_\infty^n)$ denotes the family of all probability measures on the Borel σ-algebra of the pseudo-metric space (T, d_∞^n), and $B_{d_\infty^n}(t, \varepsilon)$ denotes the ball of radius $\varepsilon > 0$ in the pseudo-metric d_∞^n with the center at $t \in T$. The random pseudo-metric d_∞^n is given by:

$$d_\infty^n(s,t) = \max_{1 \leq j \leq n} \big|f(\xi_j, s) - f(\xi_j, t)\big|$$

for $s, t \in T$ and $n \geq 1$.

4. Clearly, the conditions (3.116) and (3.117) extend under condition (3.43) to cover the unbounded case as well (see Remark 3.13). We will not pursue this in more detail here. Instead, we would like to point out that the pseudo-metric:

$$d_2^n(s,t) = \left(\frac{1}{n} \sum_{j=1}^{n} \big|f(\xi_j, s) - f(\xi_j, t)\big|^2 \right)^{1/2}$$

with $s, t \in T$, which naturally appears in (3.114) above, could also be used towards (3.116) and (3.117) instead of the pseudo-metric d_∞^n. In this context one may also like to recall Remark 3.12.

4. The Spectral Representation Theorem Approach

The purpose of the present chapter is to present a uniform law of large numbers in the wide sense stationary case which is not accessible by the previous two methods from Chapter 2 and Chapter 3. The approach relies upon the spectral representation theorem and offers conditions in terms of the orthogonal stochastic measures which are associated with the underlying dynamical system by means of this theorem. The case of bounded variation is covered (Theorem 4.3 and Theorem 4.7), while the case of unbounded variation is left as an open question.

4.1 The problem of uniform convergence in the wide sense stationary case

Let $\big(\{\xi_n(t)\}_{n \in \mathbf{Z}} \mid t \in T \big)$ be a family of (wide sense) stationary sequences of complex random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ and indexed by the set T. Then the mean-square ergodic theorem is known to be valid (see [77] p.410):

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) \rightarrow L_t \quad \text{in } L^2(P) \tag{4.1}$$

as $n \to \infty$, for all $t \in T$. The present chapter is motivated by the following question: When does the convergence in (4.1) hold uniformly over $t \in T$? In other words, when do we have:

$$\sup_{t \in T} \Big| \frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) - L_t \Big| \rightarrow 0 \quad \text{in } L^2(P) \tag{4.2}$$

as $n \to \infty$? The main purpose of this chapter is to present a solution to this problem, as well as to motivate further research in this direction (see Section 1.3). This material is taken from [69].

The main novelty of the approach towards the uniform ergodic theorem (4.2) taken in the present chapter (compare it with Chapter 2 and Chapter 3) relies upon the spectral representation theorem which is valid for (wide sense) stationary sequences under consideration. It makes possible to investigate the uniform ergodic theorem (4.2) in terms of the orthogonal stochastic measure which is associated with the underlying sequence by means of this theorem, or equivalently, in terms of the random process with orthogonal increments which corresponds to the measure.

The main result of the chapter (Theorem 4.3) states that the uniform mean-square ergodic theorem (4.2) holds as soon as the random process with orthogonal increments which is associated with the underlying sequence by means of the spectral representation theorem is of bounded variation and uniformly continuous at zero in a mean-square sense. The converse statement is also shown to be valid whenever the process is sufficiently rich. It should be mentioned that the approach of the present chapter makes no attempt to treat the case where the orthogonal stochastic measure (process with orthogonal increments) is of unbounded variation. We leave this interesting question open.

In the second part of the chapter (Section 4.3) we investigate the same problem in the continuous parameter case. Let $\big(\{X_s(t)\}_{s \in \mathbf{R}} \mid t \in T \big)$ be a family of (wide sense) stationary processes of complex random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ and indexed by the set T. Then the mean-square ergodic theorem is known to be valid (see [76] p.25):

$$\frac{1}{\tau} \int_0^{\tau} X_s(t)\, ds \rightarrow L_t \quad \text{in } L^2(P) \tag{4.3}$$

as $\tau \to \infty$, for all $t \in T$. The question under investigation is as above: When does the convergence in (4.3) hold uniformly over $t \in T$? In other words, when do we have:

$$\sup_{t \in T} \Big| \frac{1}{\tau} \int_0^\tau X_s(t)\, ds \;-\; L_t \Big| \to 0 \;\text{ in } L^2(P) \tag{4.4}$$

as $\tau \to \infty$? The main result in this context is shown to be of the same nature as the main result for sequences stated above. The same holds for the remarks following it. We will not state either of this more precisely here, but instead pass to the results in a straightforward way.

4.2 The uniform mean-square ergodic theorem (the discrete parameter case)

The aim of this section is to present a uniform mean-square ergodic theorem in the discrete parameter case. Throughout we consider a family of (*wide sense*) *stationary* sequences of complex random variables $\big(\{\xi_n(t)\}_{n \in \mathbf{Z}} \mid t \in T\big)$ defined on the probability space $(\Omega, \mathcal{F}, P)$ and indexed by the set T . Thus, we have:

$$E\big|\xi_n(t)\big|^2 < \infty \tag{4.5}$$

$$E\big(\xi_n(t)\big) = E\big(\xi_0(t)\big) \tag{4.6}$$

$$\mathrm{Cov}\big(\xi_{m+n}(t), \xi_m(t)\big) = \mathrm{Cov}\big(\xi_n(t), \xi_0(t)\big) \tag{4.7}$$

for all $n, m \in \mathbf{Z}$, and all $t \in T$. For proofs of the well-known classical results and facts which will be soon reviewed below, as well as for more information on the (wide sense) stationary property, we refer to the standard references on the subject [4], [27], [76], [77].

As a matter of convenience, we will henceforth suppose:

$$E\big(\xi_n(t)\big) = 0 \tag{4.8}$$

for all $n \in \mathbf{Z}$, and all $t \in T$. Thus the *covariance function* of $\{\xi_n(t)\}_{n \in \mathbf{Z}}$ is given by:

$$R_t(n) = E\big(\xi_n(t)\overline{\xi_0(t)}\big) \tag{4.9}$$

whenever $n \in \mathbf{Z}$ and $t \in T$.

By the *Herglotz theorem* there exists a finite measure $\mu_t = \mu_t(\Delta)$ on $\mathcal{B}(<\!-\pi, \pi\,])$ such that:

$$R_t(n) = \int_{-\pi}^{\pi} e^{in\lambda}\, \mu_t(d\lambda) \tag{4.10}$$

for $n \in \mathbf{Z}$ and $t \in T$. The measure μ_t is called the *spectral measure* of $\{\xi_n(t)\}_{n \in \mathbf{Z}}$ for $t \in T$.

The *spectral representation theorem* states that there exists an *orthogonal stochastic measure* $Z_t = Z_t(\omega, \Delta)$ on $\Omega \times \mathcal{B}(<\!-\pi, \pi\,])$ such that:

$$\xi_n(t) = \int_{-\pi}^{\pi} e^{in\lambda}\, Z_t(d\lambda) \tag{4.11}$$

for $n \in \mathbf{Z}$ and $t \in T$. The fundamental identity in this context is:

$$E\Big| \int_{-\pi}^{\pi} \varphi(\lambda)\; Z_t(d\lambda) \Big|^2 = \int_{-\pi}^{\pi} \big|\varphi(\lambda)\big|^2\, \mu_t(d\lambda) \tag{4.12}$$

whenever the function $\varphi : <\!-\pi, \pi\,] \to \mathbf{C}$ belongs to $L^2(\mu_t)$ for $t \in T$. We also have:

$$(4.13) \qquad Z_t(-\Delta) = \overline{Z_t(\Delta)}$$

for all $\Delta \in \mathcal{B}(<-\pi, \pi>)$, and all $t \in T$.

The random process defined by:

$$(4.14) \qquad Z_t(\lambda) = Z_t(<-\pi, \lambda])$$

for $\lambda \in <-\pi, \pi]$ is with *orthogonal increments* for every $t \in T$. Thus, we have:

$$(4.15) \qquad E\big|Z_t(\lambda)\big|^2 < \infty \text{ , for all } \lambda \in <-\pi, \pi]$$

$$(4.16) \qquad E\big|Z_t(\lambda_n) - Z_t(\lambda)\big|^2 \to 0 \text{ , whenever } \lambda_n \downarrow \lambda \text{ for } \lambda \in <-\pi, \pi]$$

$$(4.17) \qquad E\Big(\big(Z_t(\lambda_4) - Z_t(\lambda_3)\big)\overline{\big(Z_t(\lambda_2) - Z_t(\lambda_1)\big)}\Big) = 0$$

whenever $-\pi < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 \leq \pi$, for all $t \in T$. We will henceforth put $Z_t(-\pi) = 0$ for all $t \in T$. Moreover, we will assume below that the process $\{ Z_t(\lambda) \}_{-\pi \leq \lambda \leq \pi}$ is of *bounded variation* and *right continuous* (outside of a P-nullset) for all $t \in T$. In this case the integral:

$$(4.18) \qquad \int_{-\pi}^{\pi} \varphi(\lambda) \; Z_t(d\lambda)$$

may be well defined pointwise on Ω as the usual Riemann-Stieltjes integral for all $t \in T$. If $\psi : <\lambda_1, \lambda_2] \to \mathbf{C}$ is of bounded variation and right continuous for some $-\pi \leq \lambda_1 < \lambda_2 \leq \pi$, then *the integration by parts formula* states:

$$(4.19) \qquad \int_{\lambda_1}^{\lambda_2} \psi(\lambda-) \; Z_t(d\lambda) + \int_{\lambda_1}^{\lambda_2} Z_t(\lambda) \; \psi(d\lambda) = \psi(\lambda_2) Z_t(\lambda_2) - \psi(\lambda_1) Z_t(\lambda_1)$$

for all $t \in T$. Moreover, if we denote by $\mathrm{V}(\Phi, <\lambda_1, \lambda_2])$ the *total variation* of the function $\Phi : <\lambda_1, \lambda_2] \to \mathbf{C}$, then we have:

$$(4.20) \qquad \Big| \int_{\lambda_1}^{\lambda_2} \psi(\lambda) \; Z_t(d\lambda) \Big| \leq \sup_{\lambda_1 < \lambda \leq \lambda_2} \big|\psi(\lambda)\big| \cdot \mathrm{V}(Z_t, <\lambda_1, \lambda_2])$$

$$(4.21) \qquad \Big| \int_{\lambda_1}^{\lambda_2} Z_t(\lambda) \; \psi(d\lambda) \Big| \leq \sup_{\lambda_1 < \lambda \leq \lambda_2} \big|Z_t(\lambda)\big| \cdot \mathrm{V}(\psi, <\lambda_1, \lambda_2])$$

for all $t \in T$.

The *mean-square ergodic theorem* for $\{\xi_n(t)\}_{n \in \mathbf{Z}}$ states:

$$(4.22) \qquad \frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) \to Z_t(\{0\}) \text{ in } L^2(P)$$

as $n \to \infty$, for all $t \in T$. If moreover the process $\{ Z_t(\lambda) \}_{-\pi \leq \lambda \leq \pi}$ is of bounded variation and right continuous for all $t \in T$, then the convergence in (4.22) is P-a.s. as well. We also have:

$$(4.23) \qquad \frac{1}{n} \sum_{k=0}^{n-1} R_t(k) \to \mu_t(\{0\})$$

as $n \to \infty$, for all $t \in T$. Finally, it is easily seen that:

$$Z_t(\{0\}) = 0 \ \Leftrightarrow \ \mu_t(\{0\}) = 0 \tag{4.24}$$

whenever $t \in T$.

The main purpose of the present section is to investigate when the *uniform mean-square ergodic theorem* is valid:

$$\sup_{t \in T} \Big| \frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) - Z_t(\{0\}) \Big| \rightarrow 0 \ \text{ in } L^2(P) \tag{4.25}$$

as $n \rightarrow \infty$. We think that this problem appears worthy of consideration, and moreover to the best of our knowledge it has not been studied previously.

The main novelty of the approach towards uniform ergodic theorem taken here relies upon the spectral representation (4.11) which makes possible to investigate (4.25) in terms of the orthogonal stochastic measure $Z_t(\omega, \Delta)$ defined on $\Omega \times \mathcal{B}(<-\pi, \pi])$, or equivalently, in terms of the random process $\{ Z_t(\lambda) \}_{-\pi \leq \lambda \leq \pi}$ with orthogonal increments which corresponds to the measure by means of (4.14), where t ranges over T . In the sequel we find it convenient to restrict ourselves to the case where the process $\{ Z_t(\lambda) \}_{-\pi \leq \lambda \leq \pi}$ is of bounded variation and right continuous for $t \in T$. It is an open interesting question if the results which are obtained below under these hypotheses extend in some form to the general case.

One may observe that certain measurability problems related to (4.25) could appear (when the supremum is taken over an uncountable set). It is due to our general hypothesis on the set T . Despite this drawback we will implicitly assume measurability wherever needed. We emphasize that this simplification is not essential, and can be supported in quite a general setting by using theory of analytic spaces as explained in Paragraph 5 of Introduction. The following definition shows useful in the main theorem below.

Definition 4.1

Let $\{\xi_n(t)\}_{n \in \mathbf{Z}}$ be a (wide sense) stationary sequence of complex random variables for which the spectral representation (4.11) is valid with the process $\{ Z_t(\lambda) \}_{-\pi \leq \lambda \leq \pi}$ being of bounded variation and right continuous for $t \in T$. Then the family $\big(\{ \xi_n(t) \}_{n \in \mathbf{Z}} \mid t \in T \big)$ is said to be *variationally rich*, if for any given $-\pi \leq \lambda_1 < \lambda_2 < \lambda_3 \leq \pi$ and $t', t'' \in T$ one can find $t^* \in T$ satisfying:

$$\mathbf{V}(Z_{t'}, <\lambda_1, \lambda_2]) + \mathbf{V}(Z_{t''}, <\lambda_2, \lambda_3]) \leq \mathbf{V}(Z_{t^*}, <\lambda_1, \lambda_3]) \ . \tag{4.26}$$

It should be noted that every one point family is variationally rich. A typical non-trivial example of variationally rich family is presented in Example 4.9 below. Variationally rich families satisfy the following important property.

Lemma 4.2

Let $\big(\{\xi_n(t)\}_{n \in \mathbf{Z}} \mid t \in T \big)$ *be variationally rich, and suppose that:*

$$E\Big(\sup_{t \in T} \mathbf{V}^2\big(Z_t, <-\pi, \pi]\big) \Big) < \infty \ . \tag{4.27}$$

If $I_n = <\alpha_n, \beta_n]$ *are disjoint intervals in* $<-\pi, \pi]$ *with* $\alpha_n = \beta_{n+1}$ *for* $n \geq 1$ *, then we have:*

(2.28) $$\sup_{t\in T} \mathbf{V}(Z_t, I_n) \to 0 \quad in \quad L^2(P)$$

as $n \to \infty$.

Proof. Given $\varepsilon > 0$, choose $t_n \in T$ such that:

$$\sup_{t\in T} \mathbf{V}(Z_t, I_n) - \frac{\varepsilon}{2^n} \leq \mathbf{V}(Z_{t_n}, I_n)$$

for $n \geq 1$. Given t_n , $t_{n+1} \in T$, by (4.26) one can select $t^* \in T$ such that:

$$\mathbf{V}(Z_{t_n}, I_n) + \mathbf{V}(Z_{t_{n+1}}, I_{n+1}) \leq \mathbf{V}(Z_{t^*}, I_n \cup I_{n+1}) \leq \sup_{t\in T} \mathbf{V}(Z_t, <\!-\pi, \pi]) \ .$$

Applying the same argument to t^* and t_{n+2} , and then continuing by induction, we obtain:

$$\sum_{n=1}^{\infty} \sup_{t\in T} \mathbf{V}(Z_t, I_n) - \varepsilon \ \leq \ \sup_{t\in T} \ \mathbf{V}(Z_t, <\!-\pi, \pi]) \ .$$

Letting $\varepsilon \downarrow 0$, we get:

$$\sum_{n=1}^{\infty} \sup_{t\in T} \mathbf{V}^2(Z_t, I_n) \leq \left(\sum_{n=1}^{\infty} \sup_{t\in T} \mathbf{V}(Z_t, I_n) \right)^2 \leq \sup_{t\in T} \ \mathbf{V}^2(Z_t, <\!-\pi, \pi]) \ .$$

Taking expectation and using condition (4.27), we obtain (4.28). This completes the proof. □

We may now state the main result of this section.

Theorem 4.3

Let $\{\xi_n(t)\}_{n\in\mathbf{Z}}$ *be a (wide sense) stationary sequence of complex random variables for which the spectral representation (4.11) is valid with the process* $\{\, Z_t(\lambda) \,\}_{-\pi\leq\lambda\leq\pi}$ *being of bounded variation and right continuous for* $t \in T$ *. Suppose that the following condition is satisfied:*

(4.29) $$E\Big(\sup_{t\in T} \mathbf{V}^2\big(Z_t, <\!-\pi, \pi]\big) \Big) \ < \ \infty \ .$$

Then the uniform mean-square ergodic theorem is valid:

(4.30) $$\sup_{t\in T} \Big| \frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) - Z_t(\{0\}) \Big| \to 0 \ \ in \ L^2(P)$$

as $n \to \infty$ *, as soon as either of the following two conditions is fulfilled:*

(4.31) *There exists* $0 < \alpha < 1$ *such that:*

$$\sup_{-\frac{1}{n^\alpha} < \lambda \leq \frac{1}{n^\alpha}} E\Big(\sup_{t\in T} \big|Z_t(\lambda) - Z_t(0)\big|^2 \Big) \ = \ \mathrm{o}\big(n^{\alpha-1}\big)$$

as $n \to \infty$.

(4.32) *There exist* $0 < \alpha < 1 < \beta$ *such that:*

(i) $$\sup_{t\in T} \big|Z_t(\lambda) - Z_t(0)\big| \to 0 \ \ in \ \ P\text{-}probability$$

(ii) $$\sup_{t \in T} \mathbf{V}\Big(Z_t, \langle n^{-\beta}, n^{-\alpha}] \Big) \to 0 \ \ in \ \ P\text{-}probability$$

as $\lambda \to 0$ *and* $n \to \infty$.

Moreover, if $\big(\{\xi_n(t)\}_{n \in \mathbf{Z}} \mid t \in T \big)$ *is variationally rich, then the uniform mean-square ergodic theorem (4.30) holds if and only if we have:*

$$\sup_{t \in T} \big| Z_t(\lambda) - Z_t(0) + Z_t(0-) - Z_t(-\lambda) \big| \to 0 \ \ in \ \ P\text{-}probability \tag{4.33}$$

as $\lambda \to 0$. *In particular, if* $\big(\{\xi_n(t)\}_{n \in \mathbf{Z}} \mid t \in T \big)$ *is variationally rich, then the uniform mean-square ergodic theorem (4.30) holds whenever the process* $\{ Z_t(\lambda) \}_{-\pi \le \lambda \le \pi}$ *is uniformly continuous at zero:*

$$\sup_{t \in T} \big| Z_t(\lambda) - Z_t(0) \big| \to 0 \ \ in \ \ P\text{-}probability \tag{4.34}$$

as $\lambda \to 0$.

Proof. Let $t \in T$ and $n \ge 1$ be given and fixed. Then by (4.11) we have:

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) = \frac{1}{n} \sum_{k=0}^{n-1} \int_{-\pi}^{\pi} e^{ik\lambda} \ Z_t(d\lambda) = \int_{-\pi}^{\pi} \varphi_n(\lambda) \ Z_t(d\lambda)$$

where $\varphi_n(\lambda) = (1/n)(e^{in\lambda} - 1)/(e^{i\lambda} - 1)$ for $\lambda \ne 0$ and $\varphi_n(0) = 1$. Hence we get:

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) - Z_t(\{0\}) &= \int_{-\pi}^{\pi} \Big(\varphi_n(\lambda) - 1_{\{0\}}(\lambda) \Big) \, Z_t(d\lambda) = \int_{-\pi}^{\pi} \psi_n(\lambda) \ Z_t(d\lambda) \\ &= \int_{-\pi}^{-\delta_n} \psi_n(\lambda) \ Z_t(d\lambda) + \int_{-\delta_n}^{\delta_n} \psi_n(\lambda) \ Z_t(d\lambda) + \int_{\delta_n}^{\pi} \psi_n(\lambda) \ Z_t(d\lambda) \end{aligned} \tag{4.35}$$

for any $0 < \delta_n < \pi$, where $\psi_n(\lambda) = \varphi(\lambda)$ for $\lambda \ne 0$ and $\psi_n(0) = 0$.

We begin by showing that (4.31) is sufficient for (4.30). The proof of this fact is carried out into two steps as follows. (The first step will be of use later on as well.)

Step 1. We choose $\delta_n \downarrow 0$ in (4.35) such that:

$$\sup_{t \in T} \Big| \int_{-\pi}^{-\delta_n} \psi_n(\lambda) \ Z_t(d\lambda) \Big| \to 0 \ \text{ in } \ L^2(P) \tag{4.36}$$

$$\sup_{t \in T} \Big| \int_{\delta_n}^{\pi} \psi_n(\lambda) \ Z_t(d\lambda) \Big| \to 0 \ \text{ in } \ L^2(P) \tag{4.37}$$

as $n \to \infty$.

First consider (4.36), and note that by (4.20) we get:

$$\begin{aligned} \Big| \int_{-\pi}^{-\delta_n} \psi_n(\lambda) \ Z_t(d\lambda) \Big| &\le \sup_{-\pi < \lambda \le -\delta_n} \big| \psi_n(\lambda) \big| \cdot \mathrm{V}\big(Z_t, <\!-\pi, -\delta_n] \big) \\ &\le \frac{2}{n} \frac{1}{|\, e^{-i\delta_n} - 1 \,|} \cdot \mathrm{V}\big(Z_t, <\!-\pi, \pi] \big) \ . \end{aligned} \tag{4.38}$$

Put $\delta_n = n^{-\alpha}$ for some $\alpha > 0$, and denote $A_n = (1/n)(1/|e^{-i\delta_n} - 1|)$. Then we have:

$$A_n^2 = \frac{1}{n^2} \frac{1}{\left| e^{-i\delta_n} - 1 \right|^2} = \frac{1}{n^2} \frac{1}{2\big(1 - \cos(n^{-\alpha})\big)} \to 0 \tag{4.39}$$

as $n \to \infty$, if and only if $\alpha < 1$. Hence by (4.38) and (4.29) we see that (4.36) holds with $\delta_n = n^{-\alpha}$ for any $0 < \alpha < 1$.

Next consider (4.37), and note that by (4.20) we get:

$$\Big| \int_{\delta_n}^{\pi} \psi_n(\lambda)\ Z_t(d\lambda) \Big| \leq 2A_n \cdot \mathrm{V}\big(Z_t, <\!-\pi, \pi]\big) \tag{4.40}$$

where A_n is clearly as above. Thus by the same argument we see that (4.37) holds with $\delta_n = n^{-\alpha}$ for any $0 < \alpha < 1$. (In the sequel δ_n is always understood in this sense.)

Step 2. Here we consider the remaining term in (4.35). First notice that from the integration by parts formula (4.19) we obtain the estimate:

$$\begin{aligned}\Big| \int_{-\delta_n}^{\delta_n} \psi_n(\lambda)\ Z_t(d\lambda) \Big| \ \leq\ & \big|\psi_n(\delta_n)\big| \cdot \big|Z_t(\delta_n)\big| + \big|\psi_n(-\delta_n)\big| \cdot \big|Z_t(-\delta_n)\big| \\ & + \Big| \int_{-\delta_n}^{\delta_n} \big(Z_t(\lambda) - Z_t(0)\big)\psi_n(d\lambda) \Big| + \big|Z_t(0)\big| \cdot \big|\psi_n(\delta_n) - \psi_n(-\delta_n)\big| \ .\end{aligned}$$

Hence by Jensen's inequality we get:

$$\begin{aligned}\sup_{t\in T} \Big| \int_{-\delta_n}^{\delta_n} \psi_n(\lambda)\ Z_t(d\lambda) \Big|^2 \ \leq\ & 4\bigg(\big|\psi_n(\delta_n)\big|^2 \cdot \sup_{t\in T} \big|Z_t(\delta_n)\big|^2 + \big|\psi_n(-\delta_n)\big|^2 \cdot \sup_{t\in T} \big|Z_t(-\delta_n)\big|^2 \\ & + \mathrm{V}\big(\psi_n, <\!-\delta_n, \delta_n]\big) \cdot \int_{-\delta_n}^{\delta_n} \sup_{t\in T} \big|Z_t(\lambda) - Z_t(0)\big|^2\, \mathrm{V}(\psi_n, d\lambda) \\ & + \sup_{t\in T} \big|Z_t(0)\big|^2 \cdot \big|\psi_n(\delta_n) - \psi_n(-\delta_n)\big|^2 \bigg) \ .\end{aligned}$$

Taking expectation and using Fubini's theorem we obtain:

$$\begin{aligned}E\Big(\sup_{t\in T} \Big| \int_{-\delta_n}^{\delta_n} \psi_n(\lambda)\ Z_t(d\lambda) \Big|^2 \Big) \leq\ & 4\bigg(\big|\psi_n(\delta_n)\big|^2 \cdot E\Big(\sup_{t\in T} \big|Z_t(\delta_n)\big|^2 \Big) + \big|\psi_n(-\delta_n)\big|^2 \cdot E\Big(\sup_{t\in T} \big|Z_t(-\delta_n)\big|^2 \Big) \\ & + \sup_{-\delta_n<\lambda\leq\delta_n} E\Big(\sup_{t\in T} \big|Z_t(\lambda) - Z_t(0)\big|^2 \Big) \cdot \mathrm{V}^2\big(\psi_n, <\!-\delta_n, \delta_n]\big) \\ & + E\Big(\sup_{t\in T} \big|Z_t(0)\big|^2 \Big) \cdot \big|\psi_n(\delta_n) - \psi_n(-\delta_n)\big|^2 \bigg) \ .\end{aligned} \tag{4.41}$$

Now note that for any $-\pi < \lambda \leq \pi$ we have:

$$|Z_t(\lambda)| \leq |Z_t(\lambda) - Z_t(\lambda_t)| + |Z_t(\lambda_t)| \leq \mathrm{V}(Z_t, <\!-\pi, \pi]) + \varepsilon$$

where λ_t is chosen to be close enough to $-\pi$ to satisfy $|Z_t(\lambda_t)| \leq \varepsilon$ by right continuity. Passing to the supremum and using (4.29) hence we get:

$$E\Big(\sup_{t\in T} \sup_{-\pi<\lambda\leq\pi} \big|Z_t(\lambda)\big|^2 \Big) < \infty \ . \tag{4.42}$$

By (4.39) we have $\psi_n(\delta_n) \to 0$ and $\psi_n(-\delta_n) \to 0$, and thus by (4.42) we see that:

$$
\begin{aligned}
(4.43) \qquad & \big|\psi_n(\delta_n)\big|^2 \cdot E\Big(\sup_{t\in T} \big|Z_t(\delta_n)\big|^2\Big) + \big|\psi_n(-\delta_n)\big|^2 \cdot E\Big(\sup_{t\in T} \big|Z_t(-\delta_n)\big|^2\Big) \\
& \qquad + E\Big(\sup_{t\in T} \big|Z_t(0)\big|^2\Big) \cdot \big|\psi_n(\delta_n) - \psi_n(-\delta_n)\big|^2 \to 0
\end{aligned}
$$

as $n \to \infty$. From (4.41) we see that it remains to estimate the total variation of ψ_n on $<-\delta_n, \delta_n]$.

For this put $F_k(\lambda) = e^{ik\lambda}$ for $0 \le k \le n-1$, and notice that we have:

$$
(4.44) \qquad \mathrm{V}\big(\psi_n, <-\delta_n, \delta_n]\big) \le 1 + \mathrm{V}\big(\varphi_n, <-\delta_n, \delta_n]\big) \le 1 + \int_{-\delta_n}^{\delta_n} |\varphi_n'(\lambda)| \; d\lambda
$$

By the Cauchy-Schwarz inequality and orthogonality of F_k's on $<-\pi, \pi]$, we obtain from (4.44) the following estimate:

$$
\begin{aligned}
(4.45) \qquad \mathrm{V}\big(\psi_n, <-\delta_n, \delta_n]\big) & \le 1 + \sqrt{2\delta_n}\, \frac{1}{n} \Big(\int_{-\pi}^{\pi} \Big| \sum_{k=0}^{n-1} ikF_k(\lambda) \Big|^2 d\lambda \Big)^{1/2} \\
& = 1 + \sqrt{2\delta_n}\, \frac{1}{n} \Big(\sum_{k=1}^{n-1} k^2 \Big)^{1/2} \le 1 + \sqrt{2\delta_n}\, \frac{1}{n}\, n^{3/2} \le C n^{(1-\alpha)/2}
\end{aligned}
$$

with some constant $C > 0$. Combining (4.35), (4.36), (4.37), (4.41), (4.43) and (4.45) we complete the proof of sufficiency of (4.31) for (4.30). This fact finishes Step 2.

We proceed by showing that (4.32) is sufficient for (4.30). For this Step 1 can stay unchanged, and Step 2 is modified as follows.

Step 3. First split up the integral:

$$
(4.46) \qquad \int_{-\delta_n}^{\delta_n} \psi_n(\lambda) \; Z_t(d\lambda) = \int_{-\delta_n}^{-\eta_n} \psi_n(\lambda) \; Z_t(d\lambda) + \int_{-\eta_n}^{\eta_n} \psi_n(\lambda) \; Z_t(d\lambda) + \int_{\eta_n}^{\delta_n} \psi_n(\lambda) \; Z_t(d\lambda)
$$

for any $0 < \eta_n < \delta_n$. Put $\eta_n = n^{-\beta}$ for some $\beta > 1$. (In the sequel η_n is always understood in this sense.) Then from (4.13) and (4.20) we get by right continuity:

$$
\begin{aligned}
(4.47) \qquad \Big| \int_{-\delta_n}^{-\eta_n} \psi_n(\lambda) \; Z_t(d\lambda) \Big| & \le \mathrm{V}\big(Z_t, <-\delta_n, -\eta_n]\big) = \mathrm{V}\big(Z_t, [-\delta_n, -\eta_n]\big) \\
& = \mathrm{V}\big(Z_t, [\eta_n, \delta_n]\big) = \mathrm{V}\big(Z_t, <\eta_n, \delta_n]\big) \; .
\end{aligned}
$$

Similarly, by (4.20) we get:

$$
(4.48) \qquad \Big| \int_{\eta_n}^{\delta_n} \psi_n(\lambda) \; Z_t(d\lambda) \Big| \le \mathrm{V}\big(Z_t, <\eta_n, \delta_n]\big) \; .
$$

From (4.29), (4.47), (4.48) and (ii) in (4.32) it follows by uniform integrability that:

$$
(4.49) \qquad \sup_{t\in T} \Big| \int_{-\delta_n}^{-\eta_n} \psi_n(\lambda) \; Z_t(d\lambda) \Big| \to 0 \;\text{ in }\; L^2(P)
$$

$$
(4.50) \qquad \sup_{t\in T} \Big| \int_{\eta_n}^{\delta_n} \psi_n(\lambda) \; Z_t(d\lambda) \Big| \to 0 \;\text{ in }\; L^2(P)
$$

as $n \to \infty$. Now by (4.35), (4.36), (4.37), (4.49) and (4.50) we see that it is enough to show:

$$\sup_{t \in T} \Big| \int_{-\eta_n}^{\eta_n} \psi_n(\lambda) \; Z_t(d\lambda) \Big| \rightarrow 0 \quad \text{in} \quad L^2(P) \tag{4.51}$$

as $n \rightarrow \infty$. For this recall $\psi_n(0) = 0$. Thus from the integration by parts formula (4.19) combined with (4.21) we get:

$$\Big| \int_{-\eta_n}^{\eta_n} \psi_n(\lambda) \; Z_t(d\lambda) \Big| \leq \Big| \int_{-\eta_n}^{0-} \psi_n(\lambda) \; Z_t(d\lambda) \Big| + \Big| \int_{0}^{\eta_n} \psi_n(\lambda) \; Z_t(d\lambda) \Big| \tag{4.52}$$

$$\leq \big| \psi_n(0-) Z_t(0-) - \psi_n(-\eta_n) Z_t(-\eta_n) \big| + \Big| \int_{-\eta_n}^{0-} Z_t(\lambda) \; \psi_n(d\lambda) \Big|$$

$$+ \big| \psi_n(\eta_n) Z_t(\eta_n) - \psi_n(0+) Z_t(0) \big| + \Big| \int_{0}^{\eta_n} Z_t(\lambda) \; \psi_n(d\lambda) \Big| \leq \big| Z_t(0-) - \psi_n(-\eta_n) Z_t(-\eta_n) \big|$$

$$+ \sup_{-\eta_n < \lambda < 0} \big| Z_t(\lambda) \big| \cdot \mathrm{V}\big(\psi_n, <-\eta_n, 0>\big) + \big| \psi_n(\eta_n) Z_t(\eta_n) - Z_t(0) \big| + \sup_{0 < \lambda \leq \eta_n} \big| Z_t(\lambda) \big| \cdot \mathrm{V}\big(\psi_n, <0, \eta_n]\big) \; .$$

Next we verify that:

$$\psi_n(-\eta_n) \rightarrow 1 \quad \& \quad \psi_n(\eta_n) \rightarrow 1 \tag{4.53}$$

as $n \rightarrow \infty$. For this note that we have:

$$\big| \psi_n(\pm\eta_n) - 1 \big| = \Big| \frac{1}{n} \sum_{k=0}^{n-1} e^{\pm ik\eta_n} - 1 \Big| \leq \max_{0 \leq k \leq n-1} \big| e^{\pm ik\eta_n} - 1 \big|$$

$$= \big| e^{\pm i(n-1)\eta_n} - 1 \big| \leq \big| e^{\pm i(1/n^{\beta-1})} - 1 \big| \rightarrow 0$$

as $n \rightarrow \infty$. This proves (4.53). In addition we show that:

$$\mathrm{V}\big(\psi_n, <-\eta_n, 0>\big) \rightarrow 0 \quad \& \quad \mathrm{V}\big(\psi_n, <0, \eta_n]\big) \rightarrow 0 \tag{4.54}$$

as $n \rightarrow \infty$. For this note that we have:

$$\mathrm{V}\big(\psi_n, <0, \eta_n]\big) \leq \frac{1}{n} \sum_{k=0}^{n-1} \mathrm{V}\big(F_k, <0, \eta_n]\big) \leq \max_{0 \leq k \leq n-1} \mathrm{V}\big(F_k, <0, \eta_n]\big)$$

$$= \mathrm{V}\big(F_{n-1}, <0, \eta_n]\big) \leq \big| 1 - \cos \eta_n \big| + \big| \sin \eta_n \big| \rightarrow 0$$

as $n \rightarrow \infty$. In the same way we obtain:

$$\mathrm{V}\big(\psi_n, <-\eta_n, 0> \big) \leq \big| 1 - \cos \eta_n \big| + \big| \sin \eta_n \big| \rightarrow 0$$

as $n \rightarrow \infty$. Thus (4.54) is established. Finally, we obviously have:

$$\big| Z_t(0-) - \psi_n(-\eta_n) Z_t(-\eta_n) \big| \leq \big| Z_t(0-) - Z_t(-\eta_n) \big| + \big| 1 - \psi_n(-\eta_n) \big| \cdot \big| Z_t(-\eta_n) \big| \tag{4.55}$$

$$\big| \psi_n(\eta_n) Z_t(\eta_n) - Z_t(0) \big| \leq \big| \psi_n(\eta_n) - 1 \big| \cdot \big| Z_t(\eta_n) \big| + \big| Z_t(\eta_n) - Z_t(0) \big| \; . \tag{4.56}$$

Combining (4.32), (4.42), (4.52), (4.53), (4.54), (4.55) and (4.56) we obtain (4.51). This fact completes the proof of sufficiency of (4.32) for (4.30). Step 3 is complete.

A slight modification of Step 3 will show that (4.33) is sufficient for (4.30) whenever the family is variationally rich. This is done in the next step.

Step 4. First consider the left side in (4.52). By (4.19) and (4.21) we have:

$$\Big| \int_{-\eta_n}^{\eta_n} \psi_n(\lambda)\ Z_t(d\lambda) \Big| \leq \Big| \int_{-\eta_n}^{0-} \psi_n(\lambda)\ Z_t(d\lambda) + \int_0^{\eta_n} \psi_n(\lambda)\ Z_t(d\lambda) \Big|$$

$$\leq \big| \psi_n(0-)Z_t(0-) - \psi_n(-\eta_n)Z_t(-\eta_n) + \psi_n(\eta_n)Z_t(\eta_n) - \psi_n(0+)Z_t(0) \big|$$

$$+ \Big| \int_{-\eta_n}^{0-} Z_t(\lambda)\ \psi_n(d\lambda) \Big| + \Big| \int_0^{\eta_n} Z_t(\lambda)\ \psi_n(d\lambda) \Big| \leq \big| \psi_n(\eta_n) - 1 \big| \cdot \big| Z_t(\eta_n) \big|$$

$$+ \big| Z_t(\eta_n) - Z_t(0) + Z_t(0-) - Z_t(-\eta_n) \big| + \big| 1 - \psi_n(-\eta_n) \big| \cdot \big| Z_t(-\eta_n) \big|$$

$$+ \sup_{-\eta_n < \lambda < 0} \big| Z_t(\lambda) \big| \cdot \mathrm{V}\big(\psi_n, <-\eta_n, 0>\big) + \sup_{0 < \lambda \leq \eta_n} \big| Z_t(\lambda) \big| \cdot \mathrm{V}\big(\psi_n, <0, \eta_n]\big)\ .$$

Hence by the same arguments as in Step 3 we obtain (4.51).

Next consider the remaining terms in (4.46). By (4.47) and (4.48) we see that it suffices to show:

$$\mathrm{V}\big(Z_t, <\eta_n, \delta_n]\big) \to 0 \quad \text{in} \quad L^2(P) \tag{4.57}$$

as $n \to \infty$. We show that (4.57) holds with $\eta_n = n^{-3/2}$ and $\delta_n = n^{-1/2}$. The general case follows by the same pattern (with possibly a three intervals argument in (4.59) below).

For this put $I_n = <\eta_n, \delta_n]$, and let $p_n = (p_{n-1})^3$ for $n \geq 2$ with $p_1 = 2$. Then the intervals I_{p_n} satisfy the hypotheses of Lemma 4.2 for $n \geq 1$, and therefore we get:

$$\sup_{t \in T} \mathrm{V}\big(Z_t, I_{p_n}\big) \to 0 \quad \text{in} \quad L^2(P) \tag{4.58}$$

as $n \to \infty$. Moreover, for $p_n < q \leq p_{n+1}$ we have:

$$\mathrm{V}\big(Z_t, I_q\big) \leq \mathrm{V}\big(Z_t, I_{p_n}\big) + \mathrm{V}\big(Z_t, I_{p_{n+1}}\big)\ . \tag{4.59}$$

Thus from (4.59) we obtain (4.57), and the proof of sufficiency of (4.33) for (4.30) follows as in Step 3. This fact completes Step 4.

In the last step we prove necessity of (4.33) for (4.30) under the assumption that the family is variationally rich.

Step 5. From the integration by parts formula (4.19) we have:

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) - Z_t\big(\{0\}\big) = \int_{-\pi}^{\pi} \psi_n(\lambda)\ Z_t(d\lambda) = \int_{-\pi}^{-\delta_n} \psi_n(\lambda)\ Z_t(d\lambda) + \int_{-\delta_n}^{-\eta_n} \psi_n(\lambda)\ Z_t(d\lambda)$$

$$+ \psi_n(0-)Z_t(0-) - \psi_n(-\eta_n)Z_t(-\eta_n) + \psi_n(\eta_n)Z_t(\eta_n) - \psi_n(0+)Z_t(0)$$

$$- \int_{-\eta_n}^{0-} Z_t(\lambda)\ \psi_n(d\lambda) - \int_0^{\eta_n} Z_t(\lambda)\ \psi_n(d\lambda)$$

$$+ \int_{\eta_n}^{\delta_n} \psi_n(\lambda)\ Z_t(d\lambda) + \int_{\delta_n}^{\pi} \psi_n(\lambda)\ Z_t(d\lambda)\ .$$

Hence we easily get:

$$\big| Z_t(\eta_n) - Z_t(0) + Z_t(0-) - Z_t(-\eta_n) \big| \leq \big| \psi_n(\eta_n) - 1 \big| \cdot \big| Z_t(\eta_n) \big| + \big| 1 - \psi_n(-\eta_n) \big| \cdot \big| Z_t(-\eta_n) \big| \tag{4.60}$$

$$+ \Big| \int_{-\pi}^{\pi} \psi_n(\lambda)\ Z_t(d\lambda) \Big| + \Big| \int_{-\pi}^{-\delta_n} \psi_n(\lambda)\ Z_t(d\lambda) \Big| + \Big| \int_{-\delta_n}^{-\eta_n} \psi_n(\lambda)\ Z_t(d\lambda) \Big| + \Big| \int_{-\eta_n}^{0-} Z_t(\lambda)\ \psi_n(d\lambda) \Big| +$$

$$+\Big|\int_0^{\eta_n} Z_t(\lambda)\ \psi_n(d\lambda)\Big| + \Big|\int_{\eta_n}^{\delta_n} \psi_n(\lambda)\ Z_t(d\lambda)\Big| + \Big|\int_{\delta_n}^{\pi} \psi_n(\lambda)\ Z_t(d\lambda)\Big|\ .$$

Finally, from (4.20) and (4.21) we obtain the estimates as in Step 3 and Step 4:

$$\Big|\int_{-\eta_n}^{0-} Z_t(\lambda)\ \psi_n(d\lambda)\Big| \le \sup_{-\eta_n<\lambda<0} \big|Z_t(\lambda)\big| \cdot \mathrm{V}\big(\psi_n, <-\eta_n, 0>\big) \tag{4.61}$$

$$\Big|\int_{0}^{\eta_n} Z_t(\lambda)\ \psi_n(d\lambda)\Big| \le \sup_{0<\lambda\le\eta_n} \big|Z_t(\lambda)\big| \cdot \mathrm{V}\big(\psi_n, <0, \eta_n]\big) \tag{4.62}$$

$$\Big|\int_{-\delta_n}^{-\eta_n} \psi_n(\lambda)\ Z_t(d\lambda)\Big| \le \mathrm{V}\big(Z_t, <\eta_n, \delta_n]\big) \tag{4.63}$$

$$\Big|\int_{\eta_n}^{\delta_n} \psi_n(\lambda)\ Z_t(d\lambda)\Big| \le \mathrm{V}\big(Z_t, <\eta_n, \delta_n]\big)\ . \tag{4.64}$$

Combining (4.36), (4.37), (4.42), (4.53), (4.54), (4.57), (4.60) and (4.61)-(4.64) we complete the proof of necessity of (4.33) for (4.30). This fact finishes Step 5. The last statement of the theorem is obvious, and the proof is complete. □

Remarks 4.4

(1) Note that Theorem 4.3 reads as follows: If the convergence in (4.31) is not uniformly fast enough (but we still have it), then examine convergence of the total variation as stated in (ii) of (4.32). The characterization (4.33) with (4.34) shows that this approach is in some sense optimal.

(2) A close look into the proof shows that we have convergence P-a.s. in (4.30), as soon as we have convergence P-a.s. either in (4.31) (without the expectation and square sign, but with $(\alpha-1)/2$), or in (i) and (ii) of (4.32). Moreover, if $\big(\{\xi_n(t)\}_{n\in\mathbf{Z}} \mid t\in T\big)$ is variationally rich, then the same fact holds as for the characterization (4.33), as well as for the sufficient condition (4.34). In all these cases the condition (4.29) could be relaxed by removing the expectation and square sign. In this way we cover a *pointwise uniform ergodic theorem for (wide sense) stationary sequences*.

(3) Under (4.29) the convergence in P-probability in either (i) or (ii) of (4.32) is equivalent to the convergence in $L^2(P)$. The same fact holds for the convergence in P-probability in either (4.33) or (4.34). It follows by uniform integrability.

(4) For the condition (ii) of (4.32) note that for every fixed $t\in T$ and any $0<\alpha<1<\beta$:

$$\mathbf{V}\Big(Z_t, \langle n^{-\beta}, n^{-\alpha}]\Big) \to 0 \quad P\text{-a.s.}$$

whenever $\{\, Z_t(\lambda)\,\}_{-\pi\le\lambda\le\pi}$ is of bounded variation and right continuous (at zero), as $n\to\infty$. Note also that under (4.29) the convergence is in $L^2(P)$ as well.

(5) It is easily verified by examining the proof above that the characterization (4.33) remains valid under (4.29), whenever the property of being variationally rich is replaced with any other property implying the condition (ii) of (4.32).

(6) It remains an open interesting question if the result of Theorem 4.3 extends in some form to the case where the associated process $\{\, Z_t(\lambda)\,\}_{-\pi\le\lambda\le\pi}$ is not of bounded variation for $t\in T$.

4.3 The uniform mean-square ergodic theorem (the continuous parameter case)

The aim of this section is to present a uniform mean-square ergodic theorem in the continuous parameter case. Throughout we consider a family of (*wide sense*) *stationary* processes of complex random variables $\big(\{X_s(t)\}_{s\in\mathbf{R}} \mid t\in T\big)$ defined on the probability space $(\Omega,\mathcal{F},P)$ and indexed by the set T . Thus, we have:

$$E\big|X_s(t)\big|^2 < \infty \tag{4.65}$$

$$E\big(X_s(t)\big) = E\big(X_0(t)\big) \tag{4.66}$$

$$\mathrm{Cov}\big(X_{r+s}(t), X_r(t)\big) = \mathrm{Cov}\big(X_s(t), X_0(t)\big) \tag{4.67}$$

for all $s, r \in \mathbf{R}$, and all $t \in T$. For proofs of the well-known classical results and facts which will be soon reviewed below, we refer (as in Section 4.2) to the standard references on the subject [4], [27], [76], [77].

As a matter of convenience, we will henceforth suppose:

$$E\big(X_s(t)\big) = 0 \tag{4.68}$$

for all $s \in \mathbf{R}$, and all $t \in T$. Thus the *covariance function* of $\{X_s(t)\}_{s\in\mathbf{R}}$ is given by:

$$R_t(s) = E\big(X_s(t)\overline{X_0(t)}\big) \tag{4.69}$$

whenever $s \in \mathbf{R}$ and $t \in T$.

By the *Bochner theorem* there exists a finite measure $\mu_t = \mu_t(\Delta)$ on $\mathcal{B}(\mathbf{R})$ such that:

$$R_t(s) = \int_{-\infty}^{\infty} e^{is\lambda}\, \mu_t(d\lambda) \tag{4.70}$$

for $s \in \mathbf{R}$ and $t \in T$. The measure μ_t is called the *spectral measure* of $\{X_s(t)\}_{s\in\mathbf{R}}$ for $t \in T$.

The *spectral representation theorem* states if R_t is continuous, then there exists an *orthogonal stochastic measure* $Z_t = Z_t(\omega, \Delta)$ on $\Omega \times \mathcal{B}(\mathbf{R})$ such that:

$$X_s(t) = \int_{-\infty}^{\infty} e^{is\lambda}\, Z_t(d\lambda) \tag{4.71}$$

for $s \in \mathbf{R}$ and $t \in T$. The fundamental identity in this context is:

$$E\Big|\int_{-\infty}^{\infty} \varphi(\lambda)\; Z_t(d\lambda)\Big|^2 = \int_{-\infty}^{\infty} \big|\varphi(\lambda)\big|^2\, \mu_t(d\lambda) \tag{4.72}$$

whenever the function $\varphi :\ \mathbf{R} \to \mathbf{C}$ belongs to $L^2(\mu_t)$ for $t \in T$. We also have (4.13) which is valid for all $\Delta \in \mathcal{B}(\mathbf{R})$, and all $t \in T$.

The random process defined by:

$$Z_t(\lambda) = Z_t(<\!-\infty, \lambda]) \tag{4.73}$$

for $\lambda \in \mathbf{R}$ is with *orthogonal increments* for every $t \in T$. Thus, we have (4.15), (4.16) and (4.17) whenever $\lambda \in \mathbf{R}$ and $-\infty < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 < \infty$ for all $t \in T$. We will henceforth put $Z_t(-\infty) = 0$ for all $t \in T$. Moreover, we will assume below again that the process $\{\, Z_t(\lambda)\, \}_{\lambda\in\mathbf{R}}$ is of *bounded variation* and *right continuous* (outside of a P-nullset) for all $t \in T$. In this case the integral:

$$\int_{-\infty}^{\infty} \varphi(\lambda)\; Z_t(d\lambda) \tag{4.74}$$

may be well defined pointwise on Ω as the usual Riemann-Stieltjes integral for all $t \in T$. If $\psi : <\!\lambda_1, \lambda_2\,] \rightarrow \mathbf{C}$ is of bounded variation and right continuous for some $-\infty \leq \lambda_1 < \lambda_2 \leq \infty$, then *the integration by parts formula* (4.19) holds for all $t \in T$. Moreover, for the total variation $\mathrm{V}(\Phi, <\!\lambda_1, \lambda_2\,])$ of the function $\Phi : <\!\lambda_1, \lambda_2] \rightarrow \mathbf{C}$ we have (4.20) and (4.21) for all $t \in T$.

The *mean-square ergodic theorem* for $\{X_s(t)\}_{s \in \mathbf{R}}$ states:

(4.75) $$\frac{1}{\tau} \int_0^\tau X_s(t)\ ds \rightarrow Z_t(\{0\}) \ \text{ in } L^2(P)$$

as $\tau \rightarrow \infty$, for all $t \in T$. If moreover the process $\{\, Z_t(\lambda)\,\}_{\lambda \in \mathbf{R}}$ is of bounded variation and right continuous for all $t \in T$, then the convergence in (4.75) is P-a.s. as well. We also have:

(4.76) $$\frac{1}{\tau} \int_0^\tau R_t(s)\ ds \rightarrow \mu_t(\{0\})$$

as $\tau \rightarrow \infty$, for all $t \in T$. Finally, it is easily seen that (4.24) is valid in the present case whenever $t \in T$.

The main purpose of the present section is to investigate when the *uniform mean-square ergodic theorem* is valid:

(4.77) $$\sup_{t \in T} \Big| \frac{1}{\tau} \int_0^\tau X_s(t)\ ds - Z_t(\{0\}) \Big| \rightarrow 0 \ \text{ in } L^2(P)$$

as $\tau \rightarrow \infty$. As before, we think that this problem appears worthy of consideration, and moreover to the best of our knowledge it has not been studied previously. It turns out that the methods developed in the last section carry over to the present case without any difficulties.

The main novelty of the approach could be explained in the same way as in Section 4.2. The same remark might be also directed to the measurability problems. We will not state either of this more precisely here, but instead recall that we implicitly assume measurability wherever needed.

The definition stated in Section 4.2 extends verbatim to the present case. Again, it is shown useful in the main theorem below.

Definition 4.5

Let $\{X_s(t)\}_{s \in \mathbf{R}}$ be a (wide sense) stationary process of complex random variables for which the spectral representation (4.71) is valid with the process $\{\, Z_t(\lambda)\,\}_{\lambda \in \mathbf{R}}$ being of bounded variation and right continuous for $t \in T$. Then the family $\big(\{X_s(t)\}_{s \in \mathbf{R}} \mid t \in T\big)$ is said to be *variationally rich*, if for any given $-\infty < \lambda_1 < \lambda_2 < \lambda_3 < \infty$ and $t', t'' \in T$ one can find $t^* \in T$ satisfying:

(4.78) $$\mathbf{V}(Z_{t'}, <\!\lambda_1, \lambda_2]) + \mathbf{V}(Z_{t''}, <\!\lambda_2, \lambda_3]) \leq \mathbf{V}(Z_{t^*}, <\!\lambda_1, \lambda_3]) \ .$$

We remark again that every one point family is variationally rich. A typical non-trivial example of variationally rich family in the present case may be constructed similarly to Example 4.9 below. Variationally rich families satisfy the following important property.

Lemma 4.6

Let $\big(\{X_s(t)\}_{s \in \mathbf{R}} \mid t \in T\big)$ *be variationally rich, and suppose that:*

(4.79) $$E\Big(\sup_{t \in T} \mathbf{V}^2\big(Z_t, \mathbf{R}\big)\Big) < \infty \ .$$

If $I_n = <\alpha_n, \beta_n]$ *are disjoint intervals in* $\mathbf{R}$ *with* $\alpha_n = \beta_{n+1}$ *for* $n \geq 1$ *, then we have:*

$$\sup_{t \in T} \mathbf{V}(Z_t, I_n) \to 0 \quad in \;\; L^2(P) \tag{4.80}$$

as $n \to \infty$ *.*

Proof. The proof is exactly the same as the proof of Lemma 4.2. The only difference is that the interval $<-\pi, \pi]$ should be replaced with $\mathbf{R}$. □

We may now state the main result of this section.

Theorem 4.7

Let $\{X_s(t)\}_{s \in \mathbf{R}}$ *be a (wide sense) stationary process of complex random variables for which the spectral representation (4.71) is valid with the process* $\{Z_t(\lambda)\}_{\lambda \in \mathbf{R}}$ *being of bounded variation and right continuous for* $t \in T$ *. Suppose that the following condition is satisfied:*

$$E\Big(\sup_{t \in T} \mathbf{V}^2\big(Z_t, \mathbf{R}\big)\Big) < \infty \, . \tag{4.81}$$

Then the uniform mean-square ergodic theorem is valid:

$$\sup_{t \in T} \Big| \frac{1}{\tau} \int_0^\tau X_s(t) \; ds - Z_t(\{0\}) \Big| \to 0 \;\; in \; L^2(P) \tag{4.82}$$

as $\tau \to \infty$ *, as soon as either of the following two conditions is fulfilled:*

(4.83) *There exists* $0 < \alpha < 1$ *such that:*

$$\sup_{-\tau^{-\alpha} < \lambda \leq \tau^{-\alpha}} E\Big(\sup_{t \in T} \big|Z_t(\lambda) - Z_t(0)\big|^2\Big) = \mathrm{o}\big(\tau^{\alpha - 1}\big)$$

as $\tau \to \infty$ *.*

(4.84) *There exist* $0 < \alpha < 1 < \beta$ *such that:*

(i) $\displaystyle\sup_{t \in T} \big|Z_t(\lambda) - Z_t(0)\big| \to 0$ *in* P*-probability*

(ii) $\displaystyle\sup_{t \in T} \mathbf{V}\Big(Z_t, \big\langle \tau^{-\beta}, \tau^{-\alpha}\big]\Big) \to 0$ *in* P*-probability*

as $\lambda \to 0$ *and* $\tau \to \infty$ *.*

Moreover, if the family $\big(\{Z_s(t)\}_{s \in \mathbf{R}} \mid t \in T\big)$ *is variationally rich, then the uniform mean-square ergodic theorem (4.82) holds if and only if we have:*

$$\sup_{t \in T} \big|Z_t(\lambda) - Z_t(0) + Z_t(0-) - Z_t(-\lambda)\big| \to 0 \;\; in \;\; P\text{-probability} \tag{4.85}$$

as $\lambda \to 0$ *. In particular, if the family* $\big(\{Z_s(t)\}_{s \in \mathbf{R}} \mid t \in T\big)$ *is variationally rich, then the uniform mean-square ergodic theorem (4.82) holds whenever the process* $\{Z_t(\lambda)\}_{\lambda \in \mathbf{R}}$ *is uniformly continuous at zero:*

$$\sup_{t \in T} \big|Z_t(\lambda) - Z_t(0)\big| \to 0 \;\; in \;\; P\text{-probability} \tag{4.86}$$

as $\lambda \to 0$.

Proof. The proof may be carried out by using precisely the same method which is presented in the proof of Theorem 4.3. Thus we find it unnecessary here to provide all of the details, but instead will present only the essential points needed to make the procedure work.

Let $\tau > 0$ be fixed. We begin by noticing that a simple Riemann approximation yields:

$$\sum_{k=1}^{n} \frac{\tau}{n} \exp\Big(ik\frac{\tau}{n}\lambda\Big) \to \int_0^\tau e^{is\lambda}\; ds$$

as $n \to \infty$, for all $\lambda \in \mathbf{R}$. Since the left side above is bounded by the constant function τ which belongs to $L^2(\mu_t)$, then by (4.72) we get:

$$\int_{-\infty}^{\infty} \Big(\sum_{k=1}^{n} \frac{\tau}{n} \exp\Big(ik\frac{\tau}{n}\lambda\Big)\Big)\; Z_t(d\lambda) = \sum_{k=1}^{n} \frac{\tau}{n} X_{k\frac{\tau}{n}}(t) \to \int_{-\infty}^{\infty} \Big(\int_0^\tau e^{is\lambda}\; ds\Big)\; Z_t(d\lambda) \;\text{ in }\; L^2(P)$$

as $n \to \infty$, for all $t \in T$. Hence we get:

$$\frac{1}{\tau}\int_0^\tau X_s(t)\; ds = \int_{-\infty}^{\infty} f_\tau(\lambda)\; Z_t(d\lambda)$$

for all $t \in T$, where $f_\tau(\lambda) = (1/\tau)(e^{i\tau\lambda}-1)/(i\lambda)$ for $\lambda \neq 0$ and $f_\tau(0) = 1$. Thus we have:

$$\frac{1}{\tau}\int_0^\tau X_s(t)\; ds - Z_t(\{0\}) = \int_{-\infty}^{\infty} \Big(f_\tau(\lambda) - 1_{\{0\}}(\lambda)\Big)\; Z_t(d\lambda) = \int_{-\infty}^{\infty} g_\tau(\lambda)\; Z_t(d\lambda)$$

for all $t \in T$, where $g_\tau(\lambda) = f_\tau(\lambda)$ for $\lambda \neq 0$ and $g_\tau(0) = 0$.

Let us first reconsider Step 1. For this note that by (4.20) we get:

(4.87) $$\Big|\int_{-\pi}^{-\delta_\tau} g_\tau(\lambda)\; Z_t(d\lambda)\;\Big| \leq \sup_{-\infty<\lambda\leq-\delta_\tau} \big|g_\tau(\lambda)\big| \cdot \mathrm{V}\big(Z_t, <-\infty, -\delta_\tau]\big) \leq \frac{2}{\tau}\frac{1}{|\, i\delta_\tau\,|} \cdot \mathrm{V}\big(Z_t, \mathbf{R}\big)$$

for all $t \in T$. Thus putting $\delta_\tau = \tau^{-\alpha}$ for some $\alpha > 0$, we see that:

(4.88) $$\sup_{t\in T} \Big|\int_{-\infty}^{-\delta_\tau} g_\tau(\lambda)\; Z_t(d\lambda)\Big| \to 0 \;\text{ in }\; L^2(P)$$

as $\tau \to \infty$, as soon as we have $0 < \alpha < 1$. In exactly the same way as for (4.88) we find:

(4.89) $$\sup_{t\in T} \Big|\int_{\delta_\tau}^{\infty} g_\tau(\lambda)\; Z_t(d\lambda)\Big| \to 0 \;\text{ in }\; L^2(P)$$

as $\tau \to \infty$. Facts (4.88) and (4.89) complete Step 1.

Next reconsider Step 2. First, it is clear that:

(4.90) $$g_\tau(-\delta_\tau) \to 0 \quad \& \quad g_\tau(\delta_\tau) \to 0$$

as $\tau \to \infty$. Next, by the same arguments as in the proof of Theorem 4.3 we obtain:

(4.91) $$E\Big(\sup_{t\in T}\sup_{\lambda\in\mathbf{R}} |Z_t(\lambda)|^2\Big) < \infty \;.$$

Thus the only what remains is to estimate the total variation of g_τ on $<-\delta_\tau, \delta_\tau]$.

For this put $G_s(\lambda) = e^{is\lambda}$ for $0 \leq s \leq \tau$, and notice that we have:

$$\mathrm{V}\big(g_\tau, <-\delta_\tau, \delta_\tau]\big) \le 1 + \mathrm{V}\big(f_\tau, <-\delta_\tau, \delta_\tau]\big) \le 1 + \int_{-\delta_\tau}^{\delta_\tau} |f_\tau'(\lambda)| \; d\lambda \; . \tag{4.92}$$

Next recall that $f_\tau(\lambda) = (1/\lambda) \int_0^\tau e^{is\lambda} \; ds$ for all $\lambda \in \mathbf{R}$. Again, by the Cauchy-Schwarz inequality (with Fubini's theorem) and orthogonality of G_s's on $<-\pi, \pi]$, we obtain from (4.92):

$$\begin{aligned} \mathrm{V}\big(g_\tau, <-\delta_\tau, \delta_\tau]\big) &\le 1 + \sqrt{2\delta_\tau}\, \frac{1}{\tau} \Big(\int_{-\pi}^{\pi} \Big| \int_0^\tau is \cdot G_s(\lambda) \; ds \Big|^2 d\lambda \Big)^{1/2} \\ &= 1 + \sqrt{2\delta_\tau}\, \frac{1}{\tau} \Big(\int_0^\tau s^2 \; ds \Big)^{1/2} = 1 + \sqrt{2\delta_\tau}\, \frac{1}{\tau} \frac{\tau^{3/2}}{\sqrt{3}} \le C\tau^{(1-\alpha)/2} \end{aligned} \tag{4.93}$$

with some constant $C > 0$. Finally, by using (4.90), (4.91) and (4.93) we can complete Step 2 as in the proof of Theorem 4.3.

To complete Step 3 as in the proof of Theorem 4.3, it is just enough to verify that:

$$g_\tau(-\eta_\tau) \to 1 \;\;\&\;\; g_\tau(\eta_\tau) \to 1 \tag{4.94}$$

$$\mathrm{V}\big(g_\tau, <-\eta_\tau, 0> \big) \to 0 \;\;\&\;\; \mathrm{V}\big(g_\tau, <0, \eta_\tau]\big) \to 0 \tag{4.95}$$

as $\tau \to \infty$, where $\eta_\tau = n^{-\beta}$ for $\beta > 1$.

First consider (4.94), and note that we have:

$$\big|g_\tau(\pm\eta_\tau) - 1\big| = \Big| \frac{1}{\tau} \int_0^\tau e^{\pm is\eta_\tau} \; ds - 1 \Big| \le \sup_{0 \le s \le \tau} \big|e^{\pm is\eta_\tau} - 1\big| = \big|e^{\pm i\tau\eta_\tau} - 1\big| = \big|e^{\pm i(1/\tau^{\beta-1})} - 1\big| \to 0$$

as $\tau \to \infty$. This proves (4.94).

Next consider (4.95), and note that we have:

$$\begin{aligned} \mathrm{V}\big(g_\tau, <0, \eta_\tau]\big) &\le \frac{1}{\tau} \int_0^\tau \mathrm{V}\big(G_s, <0, \eta_\tau]\big) \; ds \le \sup_{0 \le s \le \tau} \mathrm{V}\big(G_s, <0, \eta_\tau]\big) = \mathrm{V}\big(G_\tau, <0, \eta_\tau]\big) \\ &\le \big|1 - \cos \eta_\tau\big| + \big|\sin \eta_\tau\big| \to 0 \end{aligned}$$

as $\tau \to \infty$. This proves the second part of (4.95). The first part follows by the same argument.

The rest of the proof can be carried out as was done in the proof of Theorem 4.3. □

We conclude this section by pointing out that Remarks 4.4 carry over in exactly the same form to cover the present case. The same remark may be directed to Example 4.8 and Example 4.9 below.

4.4 Uniform convergence of almost periodic sequences

In the next example we show how Theorem 4.3 applies to almost periodic sequences.

Example 4.8

Consider *an almost periodic sequence* of random variables:

$$\xi_n(t) = \sum_{k \in \mathbf{Z}} z_k(t) \, e^{i\lambda_k n} \tag{4.96}$$

for $n \in \mathbf{Z}$ and $t \in T$. In other words, for every fixed $t \in T$ we have:

(4.97) Random variables $z_i(t)$ and $z_j(t)$ are mutually orthogonal for all $i \ne j$:

$$E\big(z_i(t)\overline{z_j(t)}\big) = 0$$

(4.98) Numbers λ_k belong to $<-\pi, \pi]$ for $k \in \mathbf{Z}$, and satisfy $\lambda_i \neq \lambda_j$ whenever $i \neq j$

(4.99) The condition is satisfied:

$$\sum_{k \in \mathbf{Z}} E|z_k(t)|^2 < \infty .$$

Note that under (4.99) the series in (4.96) converges in the mean-square sense.

From (4.96) we see that the orthogonal stochastic measure is defined as follows:

$$Z_t(\Delta) = \sum_{k \in \mathbf{Z}, \lambda_k \in \Delta} z_k(t)$$

for $\Delta \in \mathcal{B}(<-\pi, \pi\,])$ and $t \in T$. The covariance function is given by:

$$R_t(n) = \sum_{k \in \mathbf{Z}} e^{i\lambda_k n} E|z_k(t)|^2$$

for $n \in \mathbf{Z}$ and $t \in T$.

In order to apply Theorem 4.3 we will henceforth assume:

$$E\left(\sup_{t \in T} \Big(\sum_{k \in \mathbf{Z}} |z_k(t)| \Big)^2 \right) < \infty . \tag{4.100}$$

Note that this condition implies:

$$E\left(\sup_{t \in T} \Big(\sum_{k \in \mathbf{Z}} |z_k(t)|^2 \Big) \right) < \infty . \tag{4.101}$$

Let $k_0 \in \mathbf{Z}$ be chosen to satisfy $\lambda_{k_0} = 0$, while otherwise set $z_{k_0}(t) \equiv 0$ for all $t \in T$. According to Theorem 4.3, the uniform mean-square ergodic theorem is valid:

$$\sup_{t \in T} \Big| \frac{1}{n} \sum_{k=0}^{n-1} \xi_k(t) - z_{k_0}(t) \Big| \to 0 \text{ in } L^2(P) \tag{4.102}$$

as $n \to \infty$, as soon as either of the following two conditions is fulfilled:

(4.103) There exists $0 < \alpha < 1$ such that:

$$\sup_{0 < \lambda \leq \frac{1}{n^\alpha}} E\left(\sup_{t \in T} \Big| \sum_{0 < \lambda_k \leq \lambda} z_k(t) \Big|^2 \right) + \sup_{-\frac{1}{n^\alpha} < \lambda < 0} E\left(\sup_{t \in T} \Big| \sum_{\lambda < \lambda_k \leq 0} z_k(t) \Big|^2 \right) = \mathrm{o}\big(n^{\alpha - 1}\big)$$

as $n \to \infty$.

(4.104) There exist $0 < \alpha < 1 < \beta$ such that:

(i) $$\sup_{t \in T} \Big| \sum_{0 < \lambda_k \leq \lambda} z_k(t) \Big| + \sup_{t \in T} \Big| \sum_{-\lambda < \lambda_k \leq 0} z_k(t) \Big| \longrightarrow 0 \quad \text{in } P\text{-probability}$$

(ii) $$\sup_{t \in T} \sum_{\frac{1}{n^\beta} < \lambda_j \leq \frac{1}{n^\alpha}} |z_j(t)| \to 0 \quad \text{in } P\text{-probability}$$

as $\lambda \downarrow 0$ and $n \to \infty$.

In particular, hence we see that if zero does not belong to the closure of the sequence $\{\lambda_k\}_{k \in \mathbf{Z}}$, then (4.102) is valid. Moreover, if the condition is fulfilled:

$$(4.105) \qquad E\Big(\sum_{k\in\mathbf{Z}} \sup_{t\in T} |z_k(t)|\Big)^2 < \infty$$

with $z_{k_0}(t) \equiv 0$ for $t \in T$, then clearly (i) and (ii) of (4.104) are satisfied, even though the condition (4.103) on the speed of convergence could possibly fail. Thus, under (4.105) we again have (4.102).

Example 4.9 (Variationally rich family)

Consider the Gaussian case in the preceding example. Thus, suppose that the almost periodic sequence (2.96) is given:

$$(4.106) \qquad \xi_n(t) = \sum_{k\in\mathbf{Z}} z_k(t)\, e^{i\lambda_k n}$$

for $n \in \mathbf{Z}$ and $t \in T$, where for every fixed $t \in T$ the random variables $z_k(t) = \sigma_k(t) \cdot g_k \sim N(0, \sigma_k^2(t))$ are independent and Gaussian with zero mean and variance $\sigma_k^2(t)$ for $k \in \mathbf{Z}$. Then (4.97) is fulfilled. We assume that (4.98) and (4.99) hold. Thus, the family $\Sigma = \big(\, \{\sigma_k^2(t)\}_{k\in\mathbf{Z}} \mid t \in T \,\big)$ satisfies the following condition:

$$(4.107) \qquad \sum_{k\in\mathbf{Z}} \sigma_k^2(t) < \infty$$

for all $t \in T$. We want to see when the family $\xi = \big(\, \{\xi_n(t)\}_{n\in\mathbf{Z}} \mid t \in T \,\big)$ is variationally rich, and this should be expressed in terms of the family Σ .

For this, take arbitrary $-\pi \le \lambda_1 < \lambda_2 < \lambda_3 \le \pi$ and t' , $t'' \in T$, and compute the left-hand side in (4.26). From the form of the orthogonal stochastic measure $Z_t = Z_t(\omega, \Delta)$ which is established in the preceding example, we see that:

$$(4.108) \qquad V(Z_t, \Delta) = \sum_{k\in\mathbf{Z},\lambda_k\in\Delta} |z_k(t)|$$

for $\Delta \in \mathcal{B}(<\!-\pi, \pi\,])$ and $t \in T$. Hence we find:

$$(4.109) \qquad \mathbf{V}(Z_{t'}, <\!\lambda_1, \lambda_2]) + \mathbf{V}(Z_{t''}, <\!\lambda_2, \lambda_3])$$
$$= \sum_{\lambda_1<\lambda_k\le\lambda_2} \big|\sigma_k(t')\big|\,|g_k| \;+\; \sum_{\lambda_2<\lambda_k\le\lambda_3} \big|\sigma_k(t'')\big|\,|g_k| \;\; .$$

Thus, in order that ξ be variationally rich, the expression in (4.109) must be dominated by:

$$(4.110) \qquad \sum_{\lambda_1<\lambda_k\le\lambda_3} \big|\sigma_k(t^*)\big|\,|g_k|$$

for some $t^* \in T$. For instance, this will be true if the family Σ satisfies the following property:

$$(4.111) \qquad \big\{\sigma_k^2(t') \vee \sigma_k^2(t'')\big\}_{k\in\mathbf{Z}} \in \Sigma$$

for all t' , $t'' \in T$. For example, if $\sigma_k(t) = t/2^{|k|}$ for $k \in \mathbf{Z}$ and t belongs to a subset T of $<0, \infty>$, the last property (4.111) is satisfied, and the family ξ is variationally rich.

Supplement: Hardy's regular convergence, uniform convergence of reversed martingales, and consistency of statistical models

The purpose of this supplement is to indicate how the concept of Hardy's regular convergence [37] plays an important role towards consistency of statistical models (see [43]) by means of *uniform convergence* of families of reversed submartingales. This material is taken from [63] and [65].

1. Let $\Pi = (\pi_\theta \mid \theta \in \Theta_0)$ be a statistical model with a sample space $(S, \mathcal{A})$, reference measure μ , and parameter set Θ_0 . In other words $(S, \mathcal{A}, \mu)$ is a measure space and π_θ is a probability measure on $(S, \mathcal{A})$ satisfying $\pi_\theta \ll \mu$ for all $\theta \in \Theta_0$. Then *the likelihood function* and *the log-likelihood function* for Π are defined as follows:

$$f(s,\theta) = \frac{d\pi_\theta}{d\mu}(s) \quad \text{and} \quad h(s,\theta) = \log f(s,\theta) \tag{S.1}$$

for all $(s,\theta) \in S \times \Theta_0$. Suppose a random phenomenon is considered that has *the unknown distribution* π belonging to Π . Then there exists $\theta_0 \in \Theta_0$ such that $\pi = \pi_{\theta_0}$ and we may define *the information function* as follows:

$$I(\theta) = \int_S f(s,\theta_0)\, h(s,\theta)\ \mu(ds) \tag{S.2}$$

for all $\theta \in \Theta_0$ for which the integral exists. Put $\beta = \sup_{\theta \in \Theta_0} I(\theta)$ and denote $M = \{ \theta \in \Theta_0 \mid I(\theta) = \beta \}$. If the following condition is satisfied:

$$\int_S f(s,\theta_0)\ |\log f(s,\theta_0)|\ \mu(ds) < \infty \tag{S.3}$$

then by the information inequality (see [43] p.32) we may conclude:

$$M = \{ \theta \in \Theta_0 \mid \pi_\theta = \pi \} \quad \text{and} \quad I(\theta) < I(\theta_0) = \beta \quad \text{for} \quad \pi_\theta \neq \pi \ . \tag{S.4}$$

Hence we see that under condition (S.3) *the problem of determining the unknown distribution* π *is equivalent to the problem of determining the set* M of all maximum points of the information function I on Θ_0 . It is easily verified that (S.3) is satisfied as soon as we have:

$$f(\,\cdot\,,\theta_0) \in L^p(\mu) \cap L^q(\mu) \tag{S.5}$$

for some $0 < p < 1 < q < \infty$. In order to determine the set M we may suppose that the observations $X_1, X_2, \ldots$ of the random phenomenon under consideration are available. In other words $\{ X_j \mid j \geq 1 \}$ is a sequence of identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in $(S, \mathcal{A})$ and the common distribution law π . If $X_1, X_2, \ldots$ are independent, then by Kolmogorov's Law 1.9 we have:

$$h_n(\,\cdot\,,\theta) = \frac{1}{n} \sum_{j=1}^n h(X_j(\,\cdot\,),\theta) \to I(\theta) \quad P\text{-a.s.} \tag{S.6}$$

as $n \to \infty$ for all $\theta \in \Theta_0$ for which the integral in (S.2) exists. Thus it may occur that under possibly additional hypotheses certain maximum points $\hat{\theta}_n$ of the map on the left side in (S.6) on Θ_0 approach the maximum points of the map on the right side in (S.6) on Θ_0 , that is the set M . This principle is, more or less explicitly, well-known and goes back to Fisher's fundamental papers [24] and [25]. A more general case of considerable interest is described as follows.

2. Some statistical models are formed by the family $\mathcal{H} = (\{h_n(\omega,\theta), \mathcal{S}_n \mid n \geq 1\} \mid \theta \in \Theta_0)$ of *reversed submartingales* defined on a probability space $(\Omega, \mathcal{F}, P)$ and indexed by a separable metric space Θ_0 . From general theory of reversed martingales (see [19]) we know that each $h_n(\theta)$ converges P-almost surely to a random variable $h_\infty(\theta)$ as $n \rightarrow \infty$, whenever $\theta \in \Theta_0$. If the tail σ-algebra $\mathcal{S}_\infty = \cap_{n=1}^\infty \mathcal{S}_n$ is degenerated, that is $P(A) \in \{0,1\}$ for all $A \in \mathcal{S}_\infty$, then $h_\infty(\theta)$ is also degenerated, that is P-almost surely equal to some constant which depends on $\theta \in \Theta_0$. In this case *the information function* associated with $\mathcal{H}$:

$$I(\theta) = a.s. \lim_{n\rightarrow\infty} h_n(\theta) = \lim_{n\rightarrow\infty} Eh_n(\theta)$$

may be well-defined for all $\theta \in \Theta_0$. The problem under consideration is *to determine the maximum points of I on Θ_0* using only information on random functions $h_n(\omega,\theta)$ for $n \geq 1$. For this reason two concepts of maximum functions may be introduced as follows.

Let $\{ \hat{\theta}_n \mid n \geq 1 \}$ be a sequence of functions from Ω into Θ , where (Θ, d) is a compact metric space containing Θ_0 . Then $\{ \hat{\theta}_n \mid n \geq 1 \}$ is called a sequence of *empirical maxima* associated with $\mathcal{H}$, if there exist a function $q : \Omega \rightarrow \mathbf{N}$ and a P-null set $N \in \mathcal{F}$ satisfying:

(S.7) $$\hat{\theta}_n(\omega) \in \Theta_0 \ , \quad \forall \omega \in \Omega \setminus N \ , \ \forall n \geq q(\omega)$$

(S.8) $$h_n(\omega, \hat{\theta}_n(\omega)) = h_n^*(\omega, \Theta_0) \ , \quad \forall \omega \in \Omega \setminus N \ , \ \forall n \geq q(\omega)$$

where $h_n^*(\omega, B) = \sup_{\theta \in B} h_n(\omega,\theta)$ for $n \geq 1$, $\omega \in \Omega$ and $B \subset \Theta_0$. The sequence $\{ \hat{\theta}_n \mid n \geq 1 \}$ is called a sequence of *approximating maxima* associated with $\mathcal{H}$, if there exist a function $q : \Omega \rightarrow \mathbf{N}$ and a P-null set $N \in \mathcal{F}$ satisfying:

(S.9) $$\hat{\theta}_n(\omega) \in \Theta_0 \ , \quad \forall \omega \in \Omega \setminus N \ , \ \forall n \geq q(\omega)$$

(S.10) $$\liminf_{n\rightarrow\infty} h_n(\omega, \hat{\theta}_n(\omega)) \geq \sup_{\theta \in \Theta_0} I(\theta) \ , \quad \forall \omega \in \Omega \setminus N \ .$$

It is easily verified that every sequence of empirical maxima is a sequence of approximating maxima. Note that although $h_n(\omega, \cdot)$ does not need to attain its maximal value on Θ_0 , and (S.8) may fail in this case, we can always find a sequence of functions $\{ \hat{\theta}_n \mid n \geq 1 \}$ satisfying:

$$h_n(\omega, \hat{\theta}_n(\omega)) \geq h_n^*(\omega, \Theta_0) - \varepsilon_n(\omega) \ , \text{ if } \ h_n^*(\omega, \Theta_0) < +\infty$$

$$h_n(\omega, \hat{\theta}_n(\omega)) \geq n \ , \text{ if } \ h_n^*(\omega, \Theta_0) = +\infty$$

for all $\omega \in \Omega$ and all $n \geq 1$, where $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Passing to the limit inferior we see that (S.10) is satisfied, so that *sequences of approximating maxima always exist.*

In order to explain the importance of the reversed-martingale assumption, we shall recall the well-known fact that any *U-statistics* (or U-process called sometimes):

$$h_n(\omega,\theta) = \frac{1}{n!} \sum_{\sigma \in \mathcal{P}_n} h(X_\sigma(\omega), \theta) \quad (\ n \geq 1 \ , \omega \in \Omega \ , \theta \in \Theta_0 \ )$$

is a reversed (sub)martingale, whenever $X = (X_1, X_2, \dots)$ is exchangeable and $Eh(X,\theta) < \infty$ for $\theta \in \Theta_0$. Here $\mathcal{P}_n$ denotes the set of all permutations of $\{ 1, 2, \dots, n \}$, and $X_\sigma =$

$(X_{\sigma_1}, \dots, X_{\sigma_n}, X_{n+1}, \dots)$.

3. The concept of consistency is introduced as follows. Define the set:

$$M = M(\mathcal{H}) = \{ \ \theta \in \Theta_0 \ | \ I(\theta) = \beta \ \}$$

where $\beta = \sup_{\theta \in \Theta_0} I(\theta)$. Let $\Gamma \subset \Theta$, then $\mathcal{H}$ is said to be *S-consistent on* Γ , if for every sequence of approximating maxima $\{ \ \hat{\theta}_n \ | \ n \geq 1 \ \}$ associated with $\mathcal{H}$ we have that $\mathcal{C}\{\hat{\theta}_n(\omega)\} \cap \Gamma \subset M$ for all $\omega \in \Omega \setminus N$, where N is a P-null set in $\mathcal{F}$. In particular $\mathcal{H}$ is said to be *S-consistent*, if it is S-consistent on Θ . Note that $\mathcal{H}$ is S-consistent on Γ if and only if, every accumulation point of any sequence of approximating maxima $\{ \ \hat{\theta}_n \ | \ n \geq 1 \ \}$ associated with $\mathcal{H}$ which belongs to Γ is a maximum point of the information function I on Θ_0 .

It can be proved (see [43] and [63]) that the *set of all possible accumulation points of all possible sequences of approximating maxima* associated with $\mathcal{H}$ equals:

$$\hat{M} = \hat{M}(\mathcal{H}) = \{ \ \theta \in \bar{\Theta}_0 \ | \ \bar{H}(\theta) \geq \beta \ \}$$

where the *upper information function* $\bar{H}$ is given by:

$$\bar{H}(\theta) = \inf_{r>0} \ \limsup_{n\to\infty} \ h_n^*(\omega, b(\theta, r)) \quad P\text{-a.s.}$$

for all $\theta \in \Theta$, and all $\omega \in \Omega$ outside some P-null set $N_\theta \in \mathcal{F}$. (We remark that all functions h_n are extended to be zero on $\Theta \setminus \Theta_0$ for $n \geq 1$.)

Hence we see that conditions implying:

$$\text{(S.11)} \qquad \limsup_{n\to\infty} \ h_n^*(\omega, b(\theta, r)) = \sup_{\xi \in b(\theta, r)} I(\xi)$$

for all $\omega \in \Omega$ outside N_θ , and all $r \in \mathbf{Q}_+$, $r \leq r_\theta$, have for a consequence:

$$\bar{H}(\theta) = \bar{I}(\theta)$$

where $\theta \in \Theta$ is a given point and $r_\theta > 0$ is a given number. Here $\bar{I}(\theta) = \lim_{r\downarrow 0} \sup_{\xi \in b(\theta, r)} I(\xi)$ denotes the upper semicontinous envelope of I for $\theta \in \Theta$. Since the set $\tilde{M} = \tilde{M}(\mathcal{H})$ is closed and contains $M(\mathcal{H})$, then $cl \ (M(\mathcal{H})) \subset \tilde{M}(\mathcal{H})$. Conversely, if $\theta \in \tilde{M}(\mathcal{H})$, then there exists a sequence $\{ \ \theta_n \ | \ n \geq 1 \ \}$ in Θ satisfying:

$$d(\theta_n, \theta) < 2^{-n} \quad \text{and} \quad I(\theta_n) \geq (\beta \wedge n) - 2^{-n}$$

for all $n \geq 1$. Thus if $\theta_n \to \theta$ with $I(\theta_n) \to \beta$ implies $I(\theta) = \beta$ for all $\theta \in \tilde{M}(\mathcal{H})$, then $\tilde{M}(\mathcal{H}) = M(\mathcal{H}) = cl(M(\mathcal{H}))$. This is for instance true if I has the closed graph, or if I is upper semicontinuous on $\tilde{M}(\mathcal{H})$. It is instructive to notice that I is always upper semicontinuous on $M(\mathcal{H})$, as well as that for every $\theta \in \tilde{M}(\mathcal{H})$ we actually have $\bar{I}(\theta) = \beta$.

Thus, the *condition (S.11) plays a central role towards consistency*. It is therefore of interest to describe this condition in more detail. This is done in terms of Hardy's regular convergence as follows. It should be noted that condition (S.11) itself is expressed by means of almost sure convergence, while Hardy's regular convergence condition presented below is in terms of means.

4. Let us recall that a double sequence of real numbers $\mathcal{A} = \{ \ a_{nk} \ | \ n, k \geq 1 \ \}$ is *convergent*

$($*in Pringsheim's sense*$)$ to the limit A , if $\forall \varepsilon > 0$, $\exists p_\varepsilon \geq 1$ such that $\forall n, k \geq p_\varepsilon$ we have $\mid A - a_{nk} \mid < \varepsilon$. In this case we shall write $A = \lim_{n,k\to\infty} a_{nk}$. According to Hardy [37] we say that a double sequence of real numbers $\mathcal{A} = \{ a_{nk} \mid n, k \geq 1 \}$ is *regularly convergent* $($*in Hardy's sense*$)$ to the limit A , if all limits:

$$\lim_{n\to\infty} a_{nk} \ , \ \lim_{k\to\infty} a_{nk} \ , \ \lim_{n,k\to\infty} a_{nk}$$

exist, for all $n, k \geq 1$, and the last one is equal to A . In this case we necessarily have:

$$\lim_{n\to\infty} \lim_{k\to\infty} a_{nk} = \lim_{k\to\infty} \lim_{n\to\infty} a_{nk} = \lim_{n,k\to\infty} a_{nk} = A \ .$$

In the proof below we shall need the following facts.

Lemma S.1

Let $\mathcal{E} = \{ E_{nk} \mid n, k \geq 1 \}$ *be a double sequence of real numbers satisfying the following two conditions:*

(S.12) $\quad E_{n1} \leq E_{n2} \leq E_{n3} \leq \ldots$, *for all* $n \geq 1$

(S.13) $\quad E_{1k} \geq E_{2k} \geq E_{3k} \geq \ldots$, *for all* $k \geq 1$.

Let $E_{n\infty} = \lim_{k\to\infty} E_{nk}$ *and* $E_{\infty k} = \lim_{n\to\infty} E_{nk}$ *for* $n, k \geq 1$. *Then the following five statements are equivalent:*

(S.14) $\quad \mathcal{E}$ *is regularly convergent* $($*in Hardy's sense*$)$

(S.15) $\quad \mathcal{E}$ *is convergent* $($*in Pringsheim's sense*$)$

(S.16) $\quad -\infty < \lim_{k\to\infty} E_{\infty k} = \lim_{n\to\infty} E_{n\infty} < +\infty$

(S.17) $\quad \forall \varepsilon > 0$, $\exists p_\varepsilon \geq 1$ *such that* $\forall n, m, k, l \geq p_\varepsilon$ *we have* $\mid E_{nk} - E_{ml} \mid < \varepsilon$

(S.18) $\quad \forall \varepsilon > 0$, $\exists p_\varepsilon \geq 1$ *such that* $E_{p_\varepsilon \infty} - E_{\infty p_\varepsilon} < \varepsilon$.

Proof. It follows by a straightforward verification. $\square$

We may now state the main result which displays a useful role of Hardy's regular convergence in establishing uniform convergence of families of reversed submartingales.

Theorem S.2

Let $(\{ X_n^i , \mathcal{F}_n \mid -\infty < n \leq -1 \} \mid i \in \mathbf{N})$ *be a countable family of reversed submartingales defined on the probability space* $(\Omega, \mathcal{F}, P)$, *and let* $X_{-\infty}^i$ *denote the* $a.s.$ *limit of* X_n^i *as* $n \to -\infty$ *for all* $i \in \mathbf{N}$. *If the following condition is satisfied:*

(S.19) $\quad \forall \varepsilon > 0$, $\exists p_\varepsilon \geq 1$ *such that* $\forall n, m, k, l \geq p_\varepsilon$ *we have:*

$$\Big| E\Big(\sup_{1\leq i\leq k} X_{-n}^i \Big) - E\Big(\sup_{1\leq j\leq l} X_{-m}^j \Big) \Big| < \varepsilon$$

then we have:

(S.20) $\quad \exists k_0 \geq 1$ *such that* $\quad -\infty < \inf_{n \leq -1} E\Big(\sup_{1 \leq i \leq k_0} X_n^i \Big) \leq \inf_{n \leq -1} E\Big(\sup_{i \in \mathbf{N}} X_n^i \Big) < +\infty$

(S.21) $\quad \sup_{i \in \mathbf{N}} X_n^i \longrightarrow \sup_{i \in \mathbf{N}} X_{-\infty}^i$ *P-a.s. and in* $L^1(P)$ *, as* $n \to -\infty$ *.*

Conversely, if we have convergence in P-probability in (S.21) together with the following condition:

(S.22) $$-\infty < \inf_{n \leq -1} E\Big(\sup_{i \in \mathbf{N}} X_n^i \Big) < +\infty$$

then (S.19) holds.

Proof. First suppose that (S.19) holds, then letting $k \to \infty$ in (S.19) and using the monotone convergence theorem we obtain:

$$\Big| E\Big(\sup_{i \in \mathbf{N}} X_{-p_\varepsilon}^i \Big) - E\Big(\sup_{1 \leq j \leq p_\varepsilon} X_{-p_\varepsilon}^j \Big) \Big| \leq \varepsilon$$

Hence we see that the last inequality in (S.20) must be satisfied, and moreover one can easily verify that for every subset M of $\mathbf{N}$ the family:

(S.23) $$\Big\{ \sup_{i \in M} X_n^i \ , \ \mathcal{F}_n \mid -\infty < n \leq p_\varepsilon \Big\}$$

forms a reversed submartingale. Thus by the reversed submartingale convergence theorem have:

(S.24) $$\sup_{i \in M} X_n^i \to X_{-\infty}^M \quad P\text{-a.s.}$$

(S.25) $$E\Big(\sup_{i \in M} X_n^i \Big) \to E\big(X_{-\infty}^M \big)$$

(S.26) $$\sup_{i \in M} X_n^i \to X_{-\infty}^M \quad \text{in} \quad L^1(P) \quad , \quad \text{if} \quad \inf_{n \leq -1} E\Big(\sup_{i \in M} X_n^i \Big) > -\infty$$

as $n \to -\infty$. Note that if M is a finite subset of $\mathbf{N}$, then $X_{-\infty}^M = \sup_{i \in M} X_{-\infty}^i$ P-a.s. Therefore letting $n \to \infty$ in (S.19) and using (S.25) we obtain:

$$\Big| E\Big(\sup_{1 \leq i \leq p_\varepsilon} X_{-\infty}^i \Big) - E\Big(\sup_{1 \leq j \leq p_\varepsilon} X_{-p_\varepsilon}^j \Big) \Big| \leq \varepsilon \ .$$

Hence $\sup_{1 \leq i \leq p_\varepsilon} X_{-\infty}^i \in L^1(P)$, and thus by (S.25) we may conclude:

$$\inf_{n \leq -1} E\Big(\sup_{1 \leq i \leq p_\varepsilon} X_n^i \Big) = E\Big(\sup_{1 \leq i \leq p_\varepsilon} X_{-\infty}^i \Big) > -\infty \ .$$

This completes the proof of (S.20). Moreover, from this by (S.26) we get:

(S.27) $$\sup_{i \in \mathbf{N}} X_n^i \to X_{-\infty}^{\mathbf{N}} \quad \text{in} \quad L^1(P)$$

as $n \to -\infty$. In order to establish (S.21) note that by (S.24) and (S.27) it is enough to show that $X_{-\infty}^{\mathbf{N}} = \sup_{i \in \mathbf{N}} X_{-\infty}^i$ P-a.s. Since the inequality $X_{-\infty}^{\mathbf{N}} \geq \sup_{i \in \mathbf{N}} X_{-\infty}^i$ P-a.s. follows straightforwardly and $X_{-\infty}^{\mathbf{N}} \in L^1(P)$, it is sufficient to show that:

(S.28) $$EX^{\mathbf{N}}_{-\infty} = E\Big(\sup_{i\in\mathbf{N}} X^i_{-\infty}\Big) \ .$$

In order to deduce (S.28) let us note that the double sequence $\mathcal{E} = \{\ E_{nk} \mid n,k \geq 1\ \}$ defined by:

(S.29) $$E_{nk} = E\Big(\sup_{1\leq i\leq k} X^i_{-n}\Big)$$

satisfies conditions (S.12) and (S.13), and moreover (S.19) is precisely (S.17). Therefore by (S.16) in Lemma S.1 and the monotone convergence theorem we may conclude:

$$EX^{\mathbf{N}}_{-\infty} = \lim_{n\to\infty}\lim_{k\to\infty} E\Big(\sup_{1\leq i\leq k} X^i_{-n}\Big) = \lim_{k\to\infty}\lim_{n\to\infty} E\Big(\sup_{1\leq i\leq k} X^i_{-n}\Big) = E\Big(\sup_{i\in\mathbf{N}} X^i_{-\infty}\Big) \ .$$

Thus (S.28) is established, and the proof of (S.21) is complete.

Next suppose that (S.22) holds and that we have convergence in P-probability in (S.21). Then by (S.23) and the second inequality in (S.22) the family $\{\ \sup_{i\in\mathbf{N}} X^i_n\ ,\ \mathcal{F}_n \mid -\infty < n \leq n_1\ \}$ is a reversed submartingale (for some $n_1 \leq -1$) which by the first inequality in (S.22) satisfies:

$$\inf_{n\leq n_1} E\Big(\sup_{i\in\mathbf{N}} X^i_n\Big) > -\infty \ .$$

Therefore (S.21) follows straightforwardly by (S.24) and (S.26). Moreover, by (S.21) and the monotone convergence theorem we may conclude:

$$-\infty < \lim_{n\to\infty}\lim_{k\to\infty} E\Big(\sup_{1\leq i\leq k} X^i_{-n}\Big) = \lim_{k\to\infty}\lim_{n\to\infty} E\Big(\sup_{1\leq i\leq k} X^i_{-n}\Big) < \infty \ .$$

Hence (S.19) follows directly by applying the implication (S.16) $\Rightarrow$ (S.17) in Lemma S.1 to the double sequence $\mathcal{E} = \{\ E_{nk} \mid n,k \geq 1\ \}$ defined by (S.29). This completes the proof. $\square$

Corollary S.3

Let $(\{\ X^i_n\ ,\ \mathcal{F}_n \mid -\infty < n \leq -1\ \} \mid i \in \mathbf{N})$ *be a countable family of reversed submartingales defined on the probability space* $(\Omega, \mathcal{F}, P)$ *, and let* $X^i_{-\infty}$ *denote the* $a.s.$ *limit of* X^i_n *as* $n \to -\infty$ *for all* $i \in \mathbf{N}$ *. Suppose that the following condition is satisfied:*

$$-\infty < \inf_{n\leq -1} E\Big(\sup_{i\in\mathbf{N}} X^i_n\Big) < +\infty \ .$$

Put $E_{nk} = E(\sup_{1\leq i\leq k} X^i_{-n})$ *for all* $n,k \geq 1$ *. Then the family:*

$$\{\ \sup_{i\in\mathbf{N}} X^i_n\ ,\ \mathcal{F}_n \mid -\infty < n \leq n_1\ \}$$

is a reversed submartingale (for some $n_1 \leq -1$ *) and* $\sup_{i\in\mathbf{N}} X^i_n$ *converges* P*-almost surely and in* $L^1(P)$ *, as* $n \to -\infty$ *. Moreover, we have:*

$$\sup_{i\in\mathbf{N}} X^i_n \longrightarrow \sup_{i\in\mathbf{N}} X^i_{-\infty} \quad P\text{-a.s. and in } L^1(P)$$

as $n \to -\infty$ *, if and only if the double sequence* $\mathcal{E} = \{\ E_{nk} \mid n,k \geq 1\ \}$ *is regularly convergent (in Hardy's sense).* $\square$

REFERENCES

[1] ANDERSEN, N. T. (1985). The calculus of non-measurable functions and sets. *Math. Inst. Aarhus, Prepr. Ser.* No. 36.

[2] ARCONES, M. A. *and* GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* 21 (1494-1542).

[3] 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).

[4] ASH, R. B. *and* GARDNER, M. F. (1975). *Topics in Stochastic Processes*. Academic Press.

[5] BERNSTEIN, S. N. (1926). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Anal.* 97 (1-59).

[6] BIRKHOFF, G. D. (1931). Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. USA* 17 (656-660).

[7] BLUM, J. R. (1955). On the convergence of empiric distribution functions. *Ann. Math. Statist.* 26 (527-529).

[8] BOLTZMANN, L. (1887). Über die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. *J. Reine Angew. Math.* 100 (201-212).

[9] BOREL, E. (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo* 27 (247-271).

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

[11] BREIMAN, L. (1968). *Probability.* Addison-Wesley.

[12] CANTELLI, F. P. (1917). Sulla probabilità come limita della frequenza. *Rend. Accad. Naz. Lincei* 26 (39-45).

[13] CANTELLI, F. P. (1933). Sulla determinazione empirica della leggi di probabilità. *Giorn. Ist. Ital. Attuari* 4 (421-424).

[14] CORNFELD, I. P., FOMIN, S. V. *and* SINAI, Ya. G. (1982). *Ergodic Theory*. Springer-Verlag, New York.

[15] DEHARDT, J. (1971). Generalizations of the Glivenko-Cantelli theorem. *Ann. Math. Statist.* 42 (2050-2055).

[16] DUNFORD, N. *and* SCHWARTZ, J. T. (1956). Convergence almost everywhere of operator averages. *J. Rat. Mech. Anal.* 5 (129-178).

[17] DUNFORD, N. *and* SCHWARTZ, J. T. (1958). *Linear Operators, Part I: General Theory*. Interscience Publ. Inc., New York.

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

[19] DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth, Inc., Belmont, California 94002.

[20] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

[21] DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Bristol.

[22] DUDLEY, R. M., GINÉ, E. *and* ZINN, J. (1991). Uniform and universal Glivenko-Cantelli classes. *J. Theoret. Probab.* 4 (435-510).

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

[24] FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London*, Ser. A, Vol. 22 (309-368).

[25] FISHER, R. A. (1925). Theory of statistical estimation. *Math. Proc. Cambridge Phil. Soc.* 22 (700-725).

[26] FURSTENBERG, H. (1981). *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press.

[27] GAENSSLER, P. *and* STUTE, W. (1979). Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Probab.* 7 (193-243).

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

[29] GARSIA, A. (1965). A simple proof of E. Hopf's maximal ergodic theorem. *J. Math. Mech.* 14 (381-382).

[30] GARSIA, A. M. (1970). Topics in Almost Everywhere Convergence. *Lectures in Advanced Mathematics* 4, Markham Publishing Company.

[31] GIHMAN, I. I. *and* SKOROHOD, A. V. (1974, 1975, 1979). *Theory of Stochastic Processes* I, II, III. Springer-Verlag, New York-Berlin.

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

[33] GINÉ, E. (1996). Empirical processes and applications: an overview. (With Discussion by J. A. Wellner). *Bernoulli* 2 (1-38).

[34] GLIVENKO, V. I. (1933). Sulla determizaione empirica della leggi di probabilita. *Giorn. Ist. Ital. Attuari* 4 (92-99).

[35] GNEDENKO, B. V. *and* KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Variables*. Addison-Wesley, Cambridge, MA.

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

[37] HARDY, G. H. (1917). On the convergence of certain multiple series. *Math. Proc. Cambridge Phil. Soc.* 19 (86-95).

[38] HARTMAN, P. *and* WINTNER, A. (1941). On the law of the iterated logarithm. *Amer. J. Math.* 63 (169-176).

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

[40] HOFFMANN-JØRGENSEN, J. (1985). The law of large numbers for non-measurable and non-separable random elements. *Astérisque* 131 (299–356).

[41] HOFFMANN-JØRGENSEN, J. (1990). Uniform convergence of martingales. *Proc. Probab. Banach Spaces VII (Oberwolfach, Germany, 1988).* Progr. Probab. Vol. 21, Birkhäuser, Boston (127-137).

[42] HOFFMANN-JØRGENSEN, J. (1990). *Stochastic Processes on Polish Spaces*. Various Publ.

Ser. Vol. 39.

[43] HOFFMANN-JØRGENSEN, J. (1992). Asymptotic likelihood theory. *Various Publ. Ser.* Vol. 40 (5-192).

[44] HOPF, E. (1937). *Ergodentheorie*. Springer, Berlin.

[45] KHINTCHINE, A. (1923). Über dyadische Brüche. *Math. Z.* 18 (109-116).

[46] KHINTCHINE, A. (1924). Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math.* 6 (9-20).

[47] KHINTCHINE, A. I. (1949). *Mathematical Foundations of Statistical Mechanics.* Dover Publications.

[48] KINGMAN, J. F. C. (1968). The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. (Ser. B)* 30 (499-510).

[49] KOLMOGOROV, A. N. (1930). Sur la loi forte des grands nombres. *C. R. Acad. Sci. Paris* 191 (910-912).

[50] KOLMOGOROV, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Math. Springer, Berlin (German). *Foundations of Theory of Probability*, Chelsea, New York, 1956 (English).

[51] KOLMOGOROV, A. N. *and* TIKHOMIROV, V. M. (1959). ε-entropy and ε-capacity of sets in functional spaces. *Uspekhi Mat. Nauk* 14 (3-86). *Amer. Math. Soc. Transl. Ser. 2* 17, 1961 (277-364).

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

[53] KRYLOFF, N. *and* BOGOLIOUBOFF, N. (1937). La théorie générale de la mesure dans son application a l'étude des systèmes dynamiques de la méchanique non linéare. *Ann. Math.* 38 (65-113).

[54] LEDOUX, M. *and* TALAGRAND, M. (1991). *Probability in Banach Spaces* (*Isoperimetry and Processes*). Springer-Verlag, Berlin, Heidelberg.

[55] LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* 15 (211-225).

[56] LUKACS, E. (1970). *Characteristic Functions.* Charles Griffin, London (Second Edition).

[57] MARUYAMA, G. (1949). The harmonic analysis of stationary stochastic processes. *Mem. Fac. Sci. Kyushu Univ. Ser. A*, Vol 4 (45-106).

[58] MOURIER, E. (1951). Lois des grands nombres et théorie ergodique. *C. R. Acad. Sci. Paris* 232 (923-925).

[59] MOURIER, E. (1953). Eléments aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré* 13 (161-244).

[60] NEUMANN, J. von (1932). Proof of the quasi-ergodic hypothesis. *Proc. Nat. Acad. Sci. USA* 18 (70-82).

[61] NOBEL, A. (1995). A counterexample concerning uniform ergodic theorems for functions. *Statist. Probab. Lett.* 24, No. 2 (165-168).

[62] PESKIR, G. (1991). Perfect measures and maps. *Math. Inst. Aarhus, Preprint Ser.* No. 26 (32 pp).

[63] PESKIR, G. (1992). Consistency of statistical models described by families of reversed

submartingales. *Math. Inst. Aarhus, Prepr. Ser.* No. 9 (28 pp). *Probab. Math. Statist.* Vol. 18, No. No. 2, 1998, (289-318).

[64] PESKIR, G. (1992). Best constants in Kahane-Khintchine inequalities in Orlicz spaces. *Math. Inst. Aarhus, Prepr. Ser.* No. 10 (42 pp). *J. Multivariate Anal.* Vol. 45, No. 2, 1993 (183-216).

[65] PESKIR, G. (1992). Uniform convergence of reversed martingales. *Math. Inst. Aarhus, Prepr. Ser.* No. 21 (27 pp). *J. Theoret. Probab.* Vol. 8, No. 2, 1995 (387-415).

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

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

[68] PESKIR, G. *and* YUKICH, J. E. (1993). Uniform ergodic theorems for dynamical systems under VC entropy conditions. *Math. Inst. Aarhus, Prepr. Ser.* No. 15 (25 pp). *Proc. Probab. Banach Spaces IX (Sandbjerg, Denmark, 1993).* Progr. Probab. Vol. 35, Birkhäuser, Boston, 1994, (104-127).

[69] PESKIR, G. (1993). The uniform mean-square ergodic theorem for wide sense stationary processes. *Math. Inst. Aarhus, Prepr. Ser.* No. 29 (19 pp). *Stochastic Anal. Appl.* Vol. 16, No. 4, 1998 (697-720).

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

[71] PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer-Verlag, Berlin.

[72] POINCARÉ, H. (1899). *Les Méthodes Nouvelles de la Mécanique Céleste I (1892), II (1893), III (1899).* Gauthier-Villars, Paris.

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

[74] POLLARD, D. (1990). Empirical Processes: Theory and Applications. *NSF-CBMS Regional Conference Series in Probability in Statistics*, Vol. 2.

[75] RÉVÉSZ, P. (1968). *The Laws of Large Numbers.* Academic Press, New York.

[76] ROZANOV, Yu. A. (1967). *Stationary Random Processes*. Holden-Day.

[77] SHIRYAYEV, A. N. (1984). *Probability*. Springer-Verlag, New York.

[78] SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell. System Tech. J.* 41 (463-501).

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

[80] STOUT, W. (1974). *Almost Sure Convergence*. Academic Press, New York.

[81] STRASSEN, V. (1966). A converse to the law of iterated logarithm. *Z. Wahrsch. Verw. Geb.* 4 (265-268).

[82] STUTE, W. *and* SCHUMANN, G. (1980). A general Glivenko-Cantelli theorem for stationary sequences of random observations. *Scand. J. Stat.* 7 (102-104).

[83] SUDAKOV, V. N. (1971). Gaussian random processes and measures of solid angles in Hilbert space. Dokl. Akad. Nauk SSSR 197 (43-45). Soviet Math. Dokl. 12 (412-415).

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

[85] TALAGRAND, M. (1996). The Glivenko-Cantelli problem, ten years later. *J. Theoret. Probab.* 9 (371-384).

[86] TUCKER, H. G. (1959). A generalization of the Glivenko-Cantelli theorem. *Ann. Math. Statist.* 30 (828-830).

[87] VALENTINE, F. A. (1964). *Convex sets*. Mc Graw-Hill, New York.

[88] VAN DER VAART, A. W. *and* WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

[89] 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).

[90] 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).

[91] WEYL, H. (1916). Über die Gleichverteilung von Zahlen mod 1. *Math. Ann.* 77 (313-352).

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

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

[94] YOSIDA, K. (1938). Mean ergodic theorem in Banach spaces. *Proc. Imp. Acad. Tokyo* 14 (292-294).

[95] YOSIDA, K. *and* KAKUTANI, S. (1939). Birkhoff's ergodic theorem and the maximal ergodic theorem. *Proc. Japan Acad.* 15 (165-168).

[96] YOSIDA, K. *and* KAKUTANI, S. (1941). Operator-theoretical treatment of Markoff's process and mean ergodic theorem. *Ann. Math.* 42 (188-228).

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

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

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