V. The Individual Ergodic Theorem

In $L^2(X, \Sigma, \mu)$, convergence in the quadratic mean (i.e. in L^2-norm) does not imply pointwise convergence, and therefore, von Neumann's ergodic theorem (IV.1) did not exactly answer the original question: For which observables f and for which states x does the time mean

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(\varphi^i(x)) \quad \text{exists?}$$

But very soon afterwards, and stimulated by von Neumann's result, G.D. Birkhoff came up with a beautiful and satisfactory answer.

V.1 Theorem (G.D. Birkhoff, 1931):
Let $(X, \Sigma, \mu; \varphi)$ be an MDS. For any $f \in L^2(X, \Sigma, \mu)$ and for almost every $x \in X$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(\varphi^i(x))$$

exists.

Even today the above theorem may not be obtained as easily as its norm-counterpart (IV.1). In addition, its modern generalizations are not as far reaching as the mean ergodic theorems contained in Lecture IV. This is due to the fact that for its formulation we need the concept of μ-a.e.-convergence, which is more strictly bound to the context of measure theory. For this reason we have to restrict our efforts to L^p-spaces, but proceed axiomatically as in Lecture IV.

V.2 Definition:
Let (X, Σ, μ) be a measure space and consider $E = L^p(X, \Sigma, \mu)$, $1 \leqslant p \leqslant \infty$. $\in \mathscr{L}(E)$ is called *individually ergodic* if for every $f \in E$ the Cesáro means $T_n f := \frac{1}{n} \sum_{i=0}^{n-1} T^i f$ converge μ-a.e. to some $\bar{f} \in E$.

Remark: The convergence of $T_n f$ in the above definition has to be understood in the following sense:
For every choice of functions g_n in the equivalence classes $\widetilde{T_n f}$, $n \in \mathbb{N}$, (see B.20) there exists a μ-null set N such that $g_n(x)$ converge for any $x \in X \backslash N$. Only in (V.D.6) we shall see how a.e.-convergence of sequences in $L^p(\mu)$ can be defined without referring to the values of representants.

There exist two main results generalizing Birkhoff's theorem, one for positive contractions on L^1, the other for the reflexive L^p-spaces. But in both cases the proof is guided by the following ideas: Prove first the a.e.-convergence of the Cesàro means T on some dense subspace of E (easy!). Then prove some "Maximal Ergodic Inequality" (difficult!), and – as an easy consequence – extend the a.e.-convergence to all of E.
Here we treat only the L^1-case and refer to App. V for the L^p-theorem.

V.3 Theorem (Hopf, 1954; Dunford-Schwartz, 1956):
Let (X, Σ, μ) be a probability space, $E = L^1(X, \Sigma, \mu)$ and $T \in \mathscr{L}(E)$. If T is positive, $T\mathbf{1} \leqslant \mathbf{1}$ and $T'\mathbf{1} \leqslant \mathbf{1}$, then T is individually ergodic.

Remark: The essential assumptions may also be stated as $\|T\|_\infty \leqslant 1$ and $\|T\|_1 \leqslant 1$ for the operator norms on $\mathscr{L}(L^\infty(\mu))$ and $\mathscr{L}(L^1(\mu))$. The proof of the above "individual ergodic theorem" will not be easy, but it is presented along the lines indicated above.

V.4 Lemma:
Under the assumptions of (V.3) there exists a dense subspace E_0 of $E = L^1(X, \Sigma, \mu)$ such that the sequence of functions $T_n f$ converges with respect to $\|\cdot\|_\infty$ for every $f \in E_0$.

Proof. By (IV.6), T is mean ergodic and therefore

$$L^1(\mu) = F \oplus \overline{(\mathrm{id} - T)L^1(\mu)} = F \oplus \overline{(\mathrm{id} - T)L^\infty(\mu)},$$

where F is the T-fixed space in $L^1(\mu)$. We take $E_0 := F \oplus (\mathrm{id} - T)L^\infty(\mu)$. The convergence is obvious for $f \in F$. But for $(\mathrm{id} - T)g$, $g \in L^\infty(\mu)$, we obtain, using (IV.3.0), the positivity of T and $T\mathbf{1} \leqslant \mathbf{1}$, the estimate

$$\begin{aligned} |T_n f| &= |(\mathrm{id} - T)T_n g| = \frac{1}{n}|(\mathrm{id} - T^n)g| \leqslant \frac{1}{n}(|g| + T^n|g|) \\ &\leqslant \frac{1}{n}(\|g\|_\infty \cdot \mathbf{1} + \|g\|_\infty \cdot T^n\mathbf{1}) \leqslant \frac{2}{n}\|g\|_\infty \cdot \mathbf{1}. \end{aligned}$$

■

V.5 Lemma (maximal ergodic lemma, Hopf, 1954):
Under the assumptions of (V.3) and for $f \in L^1(X, \Sigma, \mu)$, $n \in \mathbb{N}$, $\gamma \in \mathbb{R}_+$ we define

$$f_n^* := \sup\{T_k f : 1 \leqslant k \leqslant n\} \quad \text{and} \quad A_{n,\gamma}(f) := [f_n^* > \gamma].$$

Then

$$\gamma \cdot \mu(A_{n,\gamma}(f)) \leqslant \int_{A_{n,\gamma}(f)} f \, \mathrm{d}\mu \leqslant \|f\|.$$

Proof (Garsia, 1955):
We keep f, n and γ fixed and define

$$g := \sup\Big\{\sum_{i=0}^{k-1}(T^i f - \gamma) : 1 \leqslant k \leqslant n\Big\}.$$

First we observe that $A := A_{n,\gamma}(f) = [g > 0]$. Then

$$\begin{aligned} T(g^+) &\geqslant (Tg)^+, && \text{since } 0 \leqslant T \\ &\geqslant \sup\Big\{\Big(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma T\mathbf{1})\Big)^+ : 1 \leqslant k \leqslant n\Big\}, && \text{analogously} \\ &\geqslant \sup\Big\{\Big(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma\mathbf{1})\Big)^+ : 1 \leqslant k \leqslant n\Big\}, && \text{since } T\mathbf{1} \leqslant \mathbf{1} \\ &\geqslant \sup\Big\{\Big(\sum_{i=0}^{k-1}(T^{i+1}f - \gamma\mathbf{1})\Big)^+ : 1 \leqslant k \leqslant n-1\Big\}, \end{aligned}$$

$$= \sup\Big\{\Big(\sum_{i=0}^{k}(T^i f - \gamma\mathbf{1}) - (f-\gamma\mathbf{1})\Big)^+ : 2 \leqslant k \leqslant n\Big\},$$

$$\geqslant \sup\Big\{\sum_{i=0}^{k-1}(T^i f - \gamma\mathbf{1}) - (f-\gamma\mathbf{1}) : 1 \leqslant k \leqslant n\Big\}, \qquad \geqslant g - (f-\gamma\mathbf{1}).$$

This inequality yields

$$\mathbf{1}_A \cdot (f-\gamma\mathbf{1}) \geqslant \mathbf{1}_A \cdot g - \mathbf{1}_A \cdot T(g^+) \geqslant g^+ - T(g^+).$$

Finally the hypothesis $T'\mathbf{1} \leqslant \mathbf{1}$ implies

$$\int_A (f-\gamma\mathbf{1})\,\mathrm{d}\mu = \langle \mathbf{1}_A \cdot (f-\gamma\mathbf{1}), \mathbf{1}\rangle \geqslant \langle g^+ - T(g^+), \mathbf{1}\rangle = \langle g^+, \mathbf{1}\rangle - \langle g^+, T'\mathbf{1}\rangle \geqslant 0.$$

∎

Remarks:

1. $f^* := \sup\{T_k f : k \in \mathbb{N}\}$ is finite a.e., since $\mu[f^* > m] = \mu[\sup_{n\in\mathbb{N}} f_n^* > m] \leqslant \frac{\|f\|}{m}$ for every $m \in \mathbb{N}$, and therefore
$$\mu\Big(\bigcap_{m\in\mathbb{N}} [f^* > m]\Big) = 0 \quad \text{or} \quad \mu[f^* < \infty] = \mu\Big(\bigcup_{m\in\mathbb{N}} [f^* \leqslant m]\Big) = 1.$$
2. Observe that we didn't need the assumption $\mu(X) < \infty$ in (V.5). The essential condition was that T is defined on $L^\infty(\mu)$ and $L^1(\mu)$, and contractive for $\|\cdot\|_\infty$ and $\|\cdot\|_1$.

V.6. Proof of Theorem (V.3):
We take $0 \neq f \in L^1(\mu)$ and show that

$$h_f(x) := \limsup_{n,m\in\mathbb{N}} |T_n f(x) - T_m f(x)| = 0$$

for almost every $x \in X$. With the notation introduced above we have $h_f(x) \leqslant 2|f|^*(x)$ and $h_f(x) h_{f-f_0}(x)$ for every f_0 contained in the subspace E_0 of $\|\cdot\|_\infty$-convergence found in (V.4). By the maximal ergodic inequality (V.5) we obtain for $\gamma > 0$ the estimate

$$\mu[h_f > \gamma\|f-f_0\|] = \mu[h_{f-f_0} > \gamma\|f-f_0\|] \leqslant \mu[|f-f_0|^* > \tfrac{\gamma}{2}\|f-f_0\|]$$
$$\leqslant \frac{2\|f-f_0\|}{\gamma\|f-f_0\|} = \frac{2}{\gamma}.$$

For $\varepsilon > 0$ take $\gamma = \frac{1}{\varepsilon}$, choose $f_0 \in E_0$ such that $\|f-f_0\| < \varepsilon^2$, and conclude

$$\mu[h_f > \varepsilon] \leqslant 2\varepsilon.$$

This shows that $h_f = 0$ a.e.. ∎

Remark: The limit function $\bar{f}(x) := \lim_{n\to\infty} T_n f(x)$ is equal to Pf where P denotes the projection corresponding to the mean ergodic operator T. Therefore $\bar{f}$ is contained in $L^1(\mu)$.

Since $L^2(X,\Sigma,\mu) \subseteq L^1(X,\Sigma,\mu)$ for finite measure spaces, the Birkhoff theorem (V.1) follows immediately from (V.3) for $T = T_\varphi$. Moreover we are able to justify why “ergodicity” is the adequate “ergodic hypothesis” (compare III.D.6).

V.7 Corollary:
For an MDS $(X, \Sigma, \mu; \varphi)$ the following assertions are equivalent:

(a) φ is ergodic.

(b) For all ("observables") $f \in L^1(X, \Sigma, \mu)$ and for almost every ("state") $x \in X$ we have

$$\text{time mean } := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(\varphi^i(x)) = \int_X f \, \mathrm{d}\mu =: \text{ space mean.}$$

Proof. By (IV.7.b) the limit function $\bar{f}$ is the constant function $(\mathbf{1}\otimes\mathbf{1})f = (\int_X f \, \mathrm{d}\mu)\mathbf{1}$. ∎

V.D Discussion

V.D.1. "Equicontinuity" for a.e.-convergence:
The reader might have expected, after having proved in (V.4) a.e.-convergence on a dense subspace to finish the proof of (V.3) by a simple extension argument. For norm convergence, i.e. for the convergence induced by the norm topology, this is possible by "equicontinuity" (see B.11). But in the present context, we make the following observation.

Lemma: In general, the a.e.-convergence of sequences in $L^1(X, \Sigma, \mu)$ is not a topological convergence, i.e. there exists no topology on $L^1(X, \Sigma, \mu)$ whose convergent sequences are the a.e.-convergent sequences.

Proof. A topological convergence has the "star"-property, i.e. a sequence converges to an element f if and only if every subsequence contains a subsequence convergent to f (see Peressini [1967], p. 45). Consider $([0,1], \mathcal{B}, m)$, m the Lebesgue measure. The sequence of characteristic functions of the intervals $[0, \frac{1}{2}]$, $[\frac{1}{2}, 1]$, $[0, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$, $[\frac{1}{2}, \frac{3}{4}]$, $[\frac{3}{4}, 1]$, $[0, \frac{1}{8}], \ldots$ does not converge almost everywhere, while every subsequence contains an a.e.-convergent subsequence (see A.16) ∎

Consequently, the usual topological equicontinuity arguments are of no use in proving a.e.-convergence and are replaced by the maximal ergodic lemma (V.5) in the proof of the individual ergodic theorem. In a more general context this has already been investigated by Banach [1926] and the following "extension" result is known as "Banach's principle" (see Garsia [1970]).

Proposition: Let $(S_n)_{n\in\mathbb{N}}$ (be a sequence of bounded linear operators on $L^p(X, \Sigma, \mu)$, $1 \leqslant p < \infty$, and consider

$$S^* f(x) := \sup_{n\in\mathbb{N}} |S_n f(x)|$$

and

$$G := \{f \in L^p : S_n f \text{ converges } \mu\text{-a.e.}\}$$

If there exists a positive decreasing function

$$c : \mathbb{R}_+ \to \mathbb{R}$$

such that $\lim_{\gamma\to\infty} c(\gamma) = 0$ and

$$\mu[S^* f(x) > \gamma \|f\|] \leqslant c(\gamma)$$

for all $f \in L^p(\mu)$, $\gamma > 0$, then the subspace G is closed.

Proof. Replace $\frac{\|f\|}{\gamma}$ in the proof of (V.6) by $c(\gamma)$ ■

For an abstract treatment of this problem we refer to von Weizsäcker [1974]. See also (V.D.6).

V.D.2. Mean ergodic vs. individually ergodic:
A bounded linear operator on $L^p(X, \Sigma, \mu)$ may be mean ergodic or individually ergodic, but in general no implication is valid between the two concepts.

Example 1: The (right) shift operator

$$T : (x_n) \mapsto (0, x_1, x_2, \dots)$$

on $\ell^1(\mathbb{N}) = L^1(\mathbb{N}, \Sigma, \mu)$, where $\mu(\{n\}) = 1$ for every $n \in \mathbb{N}$, is individually ergodic, but not mean ergodic (IV.D.3).

Exercise: Transfer the above example to a finite measure space.

Example 2: On $L^2([0,1], \mathcal{B}, m)$, m Lebesgue measure, there exist operators which are not individually ergodic, but contractive hence mean ergodic (see App.V.10).

But a common consequence of the mean and individual ergodic theorem may be noted: On finite measure spaces (X, Σ, μ) the L^p-convergence and the a.e.-convergence imply the μ-stochastic convergence (see App.A.16).
Therefore

$$\lim_{n\to\infty} \mu[|T_n f(x) - \bar{f}(x)| \geqslant \varepsilon] = 0$$

for every $\varepsilon > 0$, $f \in L^p$, where $\bar{f}$ denotes the limit function of the Cesàro means $T_n f$ for a mean *or* individually ergodic operator $T \in \mathscr{L}(L^p(\mu))$.

In fact, even more is true.

Theorem (Krengel [1966]): Let (X, Σ, μ) be a finite measure space and T be a positive contraction on $L^1(\mu)$. Then the Cesàro means $T_n f$ converge stochastically for every $f \in L^1(\mu)$.

V.D.3. Strong law of large numbers (concrete example)
The strong law of large numbers "is" the individual ergodic theorem. To make this evident we have to translate it from the language of probability theory into the language of MDSs. This requires some effort and will be performed in (V.D.7). Here we content ourselves with an application of the individual ergodic theorem, i.e. the strong law of large numbers, to a concrete model. As we have seen in (II.3.ii) the Bernoulli shift $B(\frac{1}{2}, \frac{1}{2})$ is an adequate model for "coin throwing". If we take $\mathbf{1}_A$ to be the characteristic function of the rectangle

$$A = \{x = (x_n) : x_0 = 1\}$$

in $\hat{X} = \{0,1\}^{\mathbb{Z}}$, then

$$\sum_{i=0}^{n-1} \mathbf{1}_A(\tau^i x), \quad \tau \text{ the shift on } \hat{X},$$

counts the appearances of "head" in the first n performances of our "experiment" $x = (x_n)$. Since $B(\frac{1}{2}, \frac{1}{2})$ is ergodic and since $\hat{\mu}(A) = \frac{1}{2}$, the individual ergodic theorem (V.7) asserts that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_A(\tau^i x) = \frac{1}{2}$$

for a.e. $x \in \hat{X}$, i.e. the average frequency of "head" in almost every "experiment" tends to $\frac{1}{2}$.

V.D.4. Borel's theorem on normal numbers:
A number $\xi \in [0,1]$ is called normal to base 10 if in its decimal expansion

$$\xi = 0, x_1 x_2 x_3 \ldots \quad , x_i \in \{0, 1, 2, \ldots, 9\},$$

every digit appears asymptotically with frequency $\frac{1}{10}$.

Theorem (Borel, 1909): Almost every number in $[0,1]$ is normal.

Proof. First we observe that the decimal expansion is well defined except for a countable subset of $[0,1]$. Modulo these points we have a bijection from $[0,1]$ onto $\hat{X} := \{0, 1, \ldots, 9\}^{\mathbb{N}}$ which maps the Lebesgue measure onto the product measure $\hat{\mu}$ with

$$\hat{\mu}\{(x_n) \in \hat{X} : x_1 = 0\} = \cdots = \hat{\mu}\{(x_n) \in \hat{X} : x_1 = 9\} = \frac{1}{10}.$$

Consider the characteristic function χ of $\{(x_n) \in \hat{X} : x_1 = 1\}$ and the operator $T : L^1(\hat{X}, \hat{\Sigma}, \hat{\mu}) \to L^1(\hat{X}, \hat{\Sigma}, \hat{\mu})$ induced by the (left) shift

$$\tau : (x_n) \mapsto (x_{n+1}).$$

Then $\sum\limits_{i=0}^{n-1} T^i \chi(x) = \sum\limits_{i=0}^{n-1} \chi(\tau^i x)$ is the number of appearances of 1 in the first n digits of $x = (x_n)$. Since T is individually ergodic with one-dimensional fixed space, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} T^i \chi(x) = \int_X \chi \, \mathrm{d}\hat{\mu} = \frac{1}{10}$$

for almost every $x \in X$. The same is true for every other digit. ∎

V.D.5. Individually ergodic operators on C(X): It seems to be natural to adapt the question of a.e.-convergence of the Cesàro means $T_n f$ to other function spaces as well. Clearly, in the topological context and for the Banach space $C(X)$ the a.e.-convergence has to be replaced by pointwise convergence everywhere. But for bounded sequences $(f_n) \subseteq C(X)$ pointwise convergence to a continuous function is equivalent to weak convergence (see App.B.18), and by (IV.4.b) this "individual" ergodicity on $C(X)$ would not be different from mean ergodicity.

Proposition: For an operator $T \in \mathscr{L}(C(X))$ satisfying $\|T^n\| \leqslant c$ the following assertions are equivalent:

(a) For every $f \in C(X)$ the Cesàro means $T_n f$ converge pointwise to a function $f \in C(X)$.
(b) T is mean ergodic.

V.D.6. A.e.-convergence is order convergence:
While the mean ergodic theorem relies on the norm structure of $L^p(\mu)$ (and therefore generalizes to Banach spaces) there is strong evidence that the individual ergodic theorem is closely related to the order structure of $L^p(\mu)$. One reason – for others see App.V – becomes apparent in the following lemma.

Lemma: An order bounded sequence $(f_n) \subseteq L^p(X, \Sigma, \mu)$, $1 \leqslant p \leqslant \infty$, converges a.e. if and only if it is "order convergent", i.e.

$$\mathrm{o} - \overline{\lim_{n\to\infty}} f_n := \inf_{k\in\mathbb{N}} \sup_{n\geqslant k} f_n = \sup_{k\in\mathbb{N}} \inf_{n\geqslant k} =: \mathrm{o} - \underline{\lim_{n\to\infty}} f_n.$$

The proof is a simple measure-theoretical argument. It is important that the "functions" f in the order limit are elements of the order complete Banach lattice $L^p(\mu)$. In particular, "null sets" and "null functions" don't occur any more. Since the sequences $(T_n f)$ in the individual ergodic theorem are unbounded one needs a slightly more general concept. We decided not to discuss such a concept here since it seems to us that a *purely* vector lattice theoretical approach to the individual ergodic theorem has yet to prove its significance.

References: Ionescu Tulcea [1969], Peressini [1967], Yoshida [1940].

V.D.7. Strong law of large numbers (proof):
As indicated in (V.D.3) this fundamental theorem of probability theory can be obtained from the individual ergodic theorem by a translation of the probabilistic language into ergodic theory.

Theorem (Kolmogorov, 1933):
Let $(f_n)_{n\in\mathbb{N}_0}$ be a sequence of independent identically distributed integrable random variables. Then $\frac{1}{n}\sum_{i=0}^{n-1} f_i$ converge a.e. to the expected value $\mathbb{E}f_0$.

Explanation of the terminology: f is a *random variable* if there is a probability space $(\Omega, \mathscr{A}, \mathcal{P})$ such that $f : \Omega \to \mathbb{R}$ is measurable (for the Borel algebra $\mathcal{B}$ on $\mathbb{R}$). The probability measure $\mathcal{P} \circ f^{-1}$ is called the *distribution* of f, and for $A \in \mathcal{B}$ one usually writes

$$\mathcal{P}[f \in A] := p(f^{-1}(A)).$$

Two random variables f_i, f_j are *identically distributed* if they have the same distribution, i.e. $p[f_i \in A] = p[f_j \in B]$ for every $A \in \mathcal{B}$. A sequence (f_n) of random variables is called *independent* if for any finite set $J \subseteq \mathbb{N}$ and any sets $A_j \in \mathcal{B}$ we have

$$\mathcal{P}[f_j \in A_j \text{ for every } j \in J] := \mathcal{P}\Big(\bigcap_{j\in J} f_j^{-1}(A_j)\Big) = \prod_{j\in J} p(f_j^{-1}(A_j)) = \prod_{j\in J} \mathcal{P}[f_j \in A_j].$$

Finally, f is *integrable* $f \in L^1(\Omega, \mathscr{A}, \mathcal{P})$, and its *expected value* is

$$\mathbb{E}f := \int_\Omega f \, \mathrm{d}\mathcal{P}(\omega) = \int_{\mathbb{R}} t \, \mathrm{d}(\mathcal{P} \circ f^{-1})(t).$$

Proof of the Theorem. Denote by μ the distribution of (f_n), i.e.

$$\mu := \mathcal{P} \circ f_n^{-1} \quad \text{for every } n \in \mathbb{N}$$

Consider

$$\hat{X} = \mathbb{R}^{\mathbb{Z}}$$

with the product measure $\hat{\mu}$ on the product σ-algebra $\hat{\Sigma}$. With the (left) shift $\tau : \hat{X} \to \hat{X}$ we obtain an MDS $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ which is a continuous version of the Bernoulli shift on a finite set (see II.3.iii). As in (III.5.ii) we can verify that $(\hat{X}, \hat{\Sigma}, \hat{\mu}; \tau)$ is ergodic, and the individual ergodic theorem implies

$$\frac{1}{n}\sum_{i=0}^{n-1} T_\tau^i \hat{f} \xrightarrow{\text{a.e.}} \int_{\hat{X}} \hat{f}\, \mathrm{d}\hat{\mu} \quad \text{for every } \hat{f} \in L^1(\hat{X}, \hat{\Sigma}, \hat{\mu}).$$

Next, denote the projections onto the i^{th} coordinate by

$$\pi_i : \hat{X} \to \mathbb{R},$$

i.e. $\pi_i((x_n)) = x_i$. By assumption, $\pi_0 \in L^1(\hat{X}, \hat{\Sigma}, wh\mu)$ and $T_\tau^i \pi_0 = \pi_i$. Therefore

$$\frac{1}{n}\sum_{i=0}^{n-1} \pi_i \xrightarrow{\text{a.e.}} \int_{\hat{X}} \pi_0\, \mathrm{d}\hat{\mu} = \int_{\mathbb{R}} t\, \mathrm{d}\mu(t) = \mathbb{E} f_0.$$

In the final step we have to transfer the a.e.-convergence on $\hat{X}$ to the a.e.-convergence on Ω. The set of all finite products $\prod_{j \in J} g_j \circ \pi_j$ with $0 \leqslant g_j \in L^1(\mathbb{R}, \mathcal{B}, \mu)$ is total in $L^1(\hat{X}, \hat{\Sigma}, \hat{\mu})$ by construction of the product σ-algebra. On these elements we define a mapping Φ by

$$\Phi\Big(\prod_{j \in J} g_j \circ \pi_j\Big) := \prod_{j \in J} g_j \circ f_j.$$

From

$$\begin{aligned} \int_{\mathbb{R}} \Big(\prod_{j \in J} g_j \circ \pi_j \, \mathrm{d}\hat{\mu}\Big) &= \prod_{j \in J} \Big(\int_{\mathbb{R}} g_j \, \mathrm{d}\mu\Big) = \prod_{j \in J} \Big(\int_\Omega g_j \circ f_j \, \mathrm{d}\mathcal{P}\Big) \\ &= \int_\Omega \prod_{j \in J} g_j \circ f_j \, \mathrm{d}\mathcal{P} = \int_\Omega \Phi\Big(\prod_{j \in J} g_j \circ f_j\Big) \, \mathrm{d}\mathcal{P} \end{aligned}$$

it follows that Φ can be extended to a linear isometry

$$\Phi : L^1(\hat{X}, \hat{\Sigma}, \hat{\mu}) \to L^1(\Omega, \mathscr{A}, \mathcal{P}).$$

But, Φ is positive, hence preserves the order structure of the L^1-spaces and by (V.D.6) the a.e.-convergence. Therefore,

$$\frac{1}{n}\sum_{i=0}^{n-1} \Phi(\pi_i) = \frac{1}{n}\sum_{i=0}^{n-1} f_i$$

converges a.e. to $\int_\Omega \Phi(\pi_0)\, \mathrm{d}\mathcal{P} = \mathbb{E} f_0$. ∎

Remark: In the proof above we constructed a Markov shift corresponding to $p(x, A) = \mu(A)$, $x \in \mathbb{R}$, $A \in \mathcal{B}$.

References: Bauer [1968], Kolmogorov [1933], Lamperti [1977].

V.D.8. Ergodic theorems for non-positive operators:
The positivity of the operator is essential for the validity of the individual ergodic theorem. It is however possible to extend such theorems to operators which are dominated by positive operators. First we recall the basic definitions from Schaefer [1974].

Let E be an order complete Banach lattice. $T \in \mathscr{L}(E)$ is called regular if T is the difference of two positive linear operators. In that case,

$$|T| := \sup(T, -T)$$

exists and the space $\mathscr{L}^r(E)$ of all regular operators becomes a Banach lattice for the regular norm

$$\|T\|_r := \||T|\|.$$

If $E = L^1(\mu)$ or $E = L^\infty(\mu)$ then $L^r(E) = \mathscr{L}(E)$ and $\|\cdot\| = \|\cdot\|_r$ (Schaefer [1974], IV.1.5). This yields an immediate extension of (V.3).

Proposition 1: Let (X, Σ, μ) be a probability space, $E = L^1(X, \Sigma, \mu)$ and $T \in \mathscr{L}(E)$. If T is a contraction on $L^1(\mu)$ and on $L^\infty(\mu)$ then T is individually ergodic.

Proof. $|T|$ still satisfies the assumptions of (V.3), hence (V.4) and (V.5) are valid for $|T|$. But $\pm T \leqslant |T|$ implies the analogous assertion for T, hence T is individually ergodic. ∎

For $1 < p < \infty$, we have $\mathscr{L}^r(L^p) \neq \mathscr{L}(L^p)$ in general but by similar arguments we obtain from (App.V.8):

Proposition 2: Every regular contraction T, i.e. $\|T\|_r \leqslant 1$, on an L^p-space, $1 < p < \infty$ is individually ergodic.

References: Chacón- Krengel [1964], Gologan [1979], Krengel [1963], Sato [1977], Schaefer [1974].

V.D.9. A non-commutative individual ergodic theorem:
$L^\infty(X, \Sigma, \mu)$ is the prototype of a commutative W^*-algebra. Without the assumption of commutativity, every W^*-algebra can be represented as a weakly closed self-adjoint operator algebra on a Hilbert space (e.g. see Sakai [1971], 1.16.7). Since such algebras play an important role in modern mathematics and mathematical physics the following generalization of the Dunford-Schwartz individual ergodic theorem may be of some interest.

Theorem (Lance, 1976; Kümmerer, 1978):
Let $\mathscr{A}$ be a W^*-algebra and $T \in \mathscr{L}(\mathscr{A})$ a weak* continuous positive linear operator such that $T\mathbf{1} \leqslant \mathbf{1}$ and $T_*\mu \leqslant \mu$ for some faithful (= strictly positive) state μ in the predual $\mathscr{A}_*$. Then the Cesàro means $T_n x$ converge almost uniformly to $\bar{x} \in \mathscr{A}$ for every $x \in \mathscr{A}$, i.e. for every $\varepsilon > 0$ there exists a projection $p_\varepsilon \in \mathscr{A}$ such that $\mu(p_\varepsilon) < \varepsilon$ and $\|(T_n x - \bar{x})(\mathbf{1} - p_\varepsilon)\| \to 0$.

References: Conze-Dang Ngoc [1978], Kümmerer [1978], Lance [1976], Yeadon [1977].