22. Pointwise Ergodic Theorems

We now move to the statements which assert that the "time averages" converge pointwise a.e., not just in a functional-theoretical sense. The first result of this kind is due to G.D. Birkhoff which (by the date of publication) formally predates J. von Neumann's result but, as rumor has it, only because Birkhoff was the editor of the journal where J. von Neumann submitted his paper:

Theorem 22.1 (Pointwise Ergodic Theorem, Birkhoff 1931) *Let $(\mathscr{X}, \mathcal{G}, \mu, \varphi)$ be a measure-preserving system with $\mu(\mathscr{X}) < \infty$. For all $f \in L^1$ there exists $\bar f \in L^1$ with*

$$\frac{1}{n}\sum_{k=0}^{n-1} f\circ\varphi^k \underset{n\to\infty}{\longrightarrow} \bar f \qquad \mu\text{-a.e. \& in } L^1 \tag{22.1}$$

Moreover, we have $\bar f\circ\varphi = \bar f$ μ-a.e. and $\|\bar f\|_1 \leqslant \|f\|$.

The argument relies on proving another, less intuitive but nevertheless interesting, ergodic theorem which in various versions is due to N. Wiener 1939 and K. Yoshida and S. Kakutani 1939 for m.p.t.'s and Hopf 1954 for positivity-preserving contractions. (Birkhoff used a somewhat different argument and, in fact, worked only in the specific setting of smooth flows on manifolds.)

Theorem 22.2 (Maximal Ergodic Theorem, Hopf 1954) *Consider a measure space $(\mathscr{X}, \mathcal{G}, \mu)$ and let T be a positivity-preserving contraction on L^1. For $f \in L^1$ and $n \geqslant 1$ abbreviate $A_n f := \frac{1}{n}\sum_{k=0}^{n-1} T^k f$ and denote $f^\star = \sup_{n\geqslant 1} A_n f$. Then*

$$\int_{\{f^\star>0\}} f\,\mathrm{d}\mu \geqslant 0 \tag{22.2}$$

Proof. We will follow a slick argument by A. Garsia 1965. (The argument of Wiener and Yoshida-Kakutani was phrased for m.p.t. a decomposition of the measure space into pieces on which the identity is proved directly, and then assembling the pieces back together.) Introduce the following notations

$$S_n f := \sum_{k=0}^{n-1} T^k f \quad\text{and}\quad M_n f := \max_{k\leqslant n} S_k f. \tag{22.3}$$

Note that all these are in L^1 provided f is. The positive part of $M_n f$ then satisfies

$$\forall k = 1,\dots,n\colon\quad (M_n f)^+ \geqslant S_k f \tag{22.4}$$

which by the positivity-preserving property of T implies

$$\forall k = 1,\dots,n\colon\quad f + T(M_n f)^+ \geqslant f + TS_k f = S_{k+1} f \tag{22.5}$$

As $(M_n f)^+ \geqslant 0$ we also have $f + T(M_n f)^+ \geqslant f = S_1 f$ and so

$$f \geqslant M_n f - T(M_n f)^+ \tag{22.6}$$

Now let $E_n = \{M_n f > 0\}$. Then

$$\begin{aligned}\int_{E_n} f\mathrm{d}\mu &\geqslant \int_{E_n} \bigl[M_n f - T(M_n f)^+\bigr]\mathrm{d}\mu \\ &\geqslant \int_{E_n} \bigl[(M_n f)^+ - T(M_n f)^+\bigr]\mathrm{d}\mu \\ &\geqslant \int_{\Omega} (M_n f)^+\mathrm{d}\mu - \int_{\Omega} T(M_n f)^+\mathrm{d}\mu\end{aligned} \tag{22.7}$$

But $h \geqslant 0$ and $h \in L^1$ imply $\int Th\mathrm{d}\mu = \|Th\| \leqslant \|h\| = \int h\mathrm{d}\mu$ and so the right-hand side of (22.7) is larger than zero. We conclude

$$\forall n \geqslant 1\colon \quad \int_{E_n} f\mathrm{d}\mu \geqslant 0 \tag{22.8}$$

But E_n increases to $\{f^\star > 0\}$ as $n \to \infty$ and so we get (22.2) by the Dominated Convergence Theorem. □

We now process the terse-looking inequality (22.2) into a very familiar form:

Corollary 22.3 (Maximal inequality) *Suppose $A \in \mathcal{F}$ be φ-invariant, i.e., $\varphi^{-1}(A) = A$, and obeys $\mu(A) < \infty$. Given $f \in L^1$ let $f^\star := \sup_{n\geqslant 1} \frac{1}{n}\sum_{k=0}^{n-1} f \circ \varphi^k$. Then*

$$\forall \alpha \in \mathbb{R}\colon \quad \alpha\mu\bigl(A \cap \{f^\star > \alpha\}\bigr) \leqslant \int_{A\cap\{f^\star>\alpha\}} f\mathrm{d}\mu \tag{22.9}$$

Proof. Let $g := (f-\alpha)1_A$. Then $\mu(A) < \infty$ implies $g \in L^1$ and the φ-invariance of A gives $g^\star = (f^\star - \alpha)1_A$. Hence $\{g^\star > 0\} = A \cap \{f^\star > \alpha\}$ a.e. and so

$$0 \leqslant \int_{\{g^\star>0\}} g\mathrm{d}\mu = \int_{A\cap\{f^\star>\alpha\}} (f-\alpha)\,\mathrm{d}\mu = -\alpha\mu\bigl(A \cap \{f^\star > \alpha\}\bigr) + \int_{A\cap\{f^\star>\alpha\}} f\mathrm{d}\mu \tag{22.10}$$

where the last step follows again by $\mu(A) < \infty$. □

We note that the requirement that $\mu(A) < \infty$ is what will eventually force us to take $\mu(\mathscr{X}) < \infty$ in the proof below. As we will show later, this is not necessary for a.e. convergence and integrability of the limit, but matters for the rest of the statement. Our proof below would work if $\mathscr{X}$ can be partitioned into countably many finite-measure invariant sets which, of course, is a fairly unrealistic requirement.

Proof of the Pointwise Ergodic Theorem. Assume $\mu(\mathscr{X}) < \infty$, pick $\alpha < \beta$ and consider the set

$$E_{\alpha\beta} := \bigl\{\liminf_{n\to\infty} A_n f < \alpha < \beta < \limsup_{n\to\infty} A_n f\bigr\}. \tag{22.11}$$

This set if φ-invariant because the limsup and liminf are, i.e., $\varphi^{-1}(E_{\alpha\beta}) = E_{\alpha\beta}$. Since $E_{\alpha\beta} \subseteq \{f^\star > \beta\}$, from (22.9) we get

$$\beta\mu(E_{\alpha\beta}) \leqslant \int_{E_{\alpha\beta}} f\mathrm{d}\mu \tag{22.12}$$

On the other hand, $E_{\alpha\beta} \subseteq \{(-f)^* > -\alpha\}$ and thus also

$$(-\alpha)\mu(E_{\alpha\beta}) \leqslant -\int_{E_{\alpha\beta}} f \mathrm{d}\mu \tag{22.13}$$

Combining these we conclude

$$\beta\mu(E_{\alpha\beta}) \leqslant \int_{E_{\alpha\beta}} f \mathrm{d}\mu \leqslant \alpha\mu(E_{\alpha\beta}). \tag{22.14}$$

But $\alpha < \beta$ and $\mu(E_{\alpha\beta}) < \infty$, so this is possible only if $\mu(E_{\alpha\beta}) = 0$. It follows that $A_n f$ converges pointwise almost everywhere to $\bar{f} := \limsup_{n\to\infty} A_n f$.

The definition ensures that $\bar{f}$ is φ-invariant. Fatou's lemma along with $|A_n f| \leqslant A_n|f|$ implies $\|\bar{f}\|_1 \leqslant \|f\|$. To show L^1-convergence, pick a non-negative $f \in L^1$ and let g be a bounded function with $0 < g < f$. Then

$$\begin{aligned} \|A_n f - \bar{f}\|_1 &\leqslant \|A_n(f-g)\|_1 + \|A_n g - \bar{g}\|_1 + \|\bar{f} - \bar{g}\|_1 \\ &\leqslant 2\|f-g\|_1 + \|A_n g - \bar{g}\|_1, \end{aligned} \tag{22.15}$$

where we used that A_n is a contraction to derive the second inequality. For g bounded we have $A_n g \to \bar{g}$ in L^1 by the Bounded Convergence Theorem. Then let $g \to f$ in L^1. $\square$

The above is an argument we used in the proof of Theorem 15.1 for Lévy martingales. Here is another proof that is more in line with what we did for the Mean Ergodic Theorems earlier:

Another proof of the Pointwise Ergodic Theorem. Assume $\mu(\mathscr{X}) < \infty$. The maximal inequality in Corollary 22.3 applied to any $f \in L^1$ then gives

$$\forall \lambda > 0\colon\ \mu\big(|f|^\star > \lambda\big) \leqslant \frac{1}{\lambda}\|f\|_1 \tag{22.16}$$

As argued before, $\lim_{n\to\infty} A_n f$ exists for all $f \in \mathscr{L}$ where

$$\mathscr{L} = \{g + h - Th\colon h, g \in L^1,\ g = g \circ \varphi\} \tag{22.17}$$

Since $\mathscr{L}$ was shown to be dense in L^2, which for finite μ is dense in L^1, for each $f \in L^1$ and each $k \geqslant 1$ there exists $f_k \in \mathscr{L}$ such that $\|f - f_k\|_1 < 4^{-k}$. But then (22.16) gives

$$\mu\big(|f - f_k|^\star > 2^{-k}\big) \leqslant 2^{-k} \tag{22.18}$$

Using the Borel-Cantelli reasoning we have

$$\mu\big(|f - f_k|^\star > 2^{-k} \text{ i.o.}\big) = 0 \tag{22.19}$$

On the full-measure set $\Omega_0 := \{|f - f_k|^\star > 2^{-k} \text{ i.o.}\}^{\mathrm{c}}$ we have $|A_n f - A_n f_k| \leqslant 2^{-k}$ once k is sufficiently large and so

$$\lim_{n\to\infty} A_n f_k - 2^{-k} \leqslant \liminf_{n\to\infty} A_n f \leqslant \limsup_{n\to\infty} A_n \leqslant \lim_{n\to\infty} A_n f_k + 2^{-k} \tag{22.20}$$

i.e.,

$$\limsup_{n\to\infty} A_n - \liminf_{n\to\infty} A_n \leqslant \frac{2}{k} \tag{22.21}$$

In short, $\lim_{n\to\infty} A_n f$ exists and (by (22.20)) is finite on Ω_0. The remainder of the proof is the same and so we will not repeat it. $\square$

For the case of the m.p.s. associated with i.i.d. random variables, the above gives

$$\{X_k\}_{k\geqslant 1} \text{ i.i.d. } \wedge\ X_1 \in L^1 \ \Rightarrow\ \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k \text{ exists a.s. \& in } L^1 \tag{22.22}$$

The Hewitt-Savage Zero-One Law (Theorem 18.6) then implies that the limit must be constant and so equal to EX_1. Hence, Theorem 22.1 gives us yet another way to prove the SLLN this time using an argument that generalizes rather dramatically to other contexts.