23. More on pointwise convergence

We will now generalize the above ideas further to give statements and proofs of a few other ergodic theorems that are interesting both for the theory and its applications.

23.1 Spatial Ergodic Theorem.

We start with a multi-dimensional version of Birkhoff's theorem. Our proof will also extend the relevant parts of Birkhoff's theorem to infinite measure spaces. Given a set $\Lambda \subseteq \mathbb{Z}^d$, denote its *edge boundary* as

$$\partial\Lambda := \bigl\{(x,y) \in \mathbb{Z}^d \times \mathbb{Z}^d \colon x \in \Lambda \wedge y \notin \Lambda \wedge \|x-y\|_\infty = 1\bigr\} \tag{23.1}$$

The statement is as follows:

Theorem 23.1 (Spatial Ergodic Theorem) *Let $(\mathscr{X}, \mathcal{G}, \mu)$ be a measure space with $\mu(\mathscr{X})$ finite or infinite. Given an integer $d \geqslant 1$, let $\varphi_1, \dots, \varphi_d$ be measure-preserving transformations on $(\mathscr{X}, \mathcal{G}, \mu)$ that commute, i.e.,*

$$\forall i,j = 1,\dots,d\colon\ \varphi_i \circ \varphi_j = \varphi_j \circ \varphi_i \tag{23.2}$$

For all $f \in L^1$ there exists $\bar f \in L^1$ such that for all sets $\Lambda_n \subseteq \mathbb{Z}^d$ with

$$\forall n \geqslant 1\colon\ \varnothing \neq \Lambda_n \subseteq [0,n)^d \tag{23.3}$$

and

$$\liminf_{n\to\infty} \frac{|\Lambda_n|}{n^d} > 0 \ \wedge\ \lim_{n\to\infty} \frac{|\partial\Lambda_n|}{n^d} = 0 \tag{23.4}$$

we have

$$\frac{1}{|\Lambda_n|} \sum_{x\in\Lambda_n} f\circ\varphi_x \underset{n\to\infty}{\longrightarrow} \bar f \quad \text{a.e.} \tag{23.5}$$

where $\varphi_{(x_1,\dots,x_n)} := \varphi_1^{x_1} \circ \dots \circ \varphi_d^{x_d}$. If $\mu(\mathscr{X}) < \infty$ then the convergence is in L^1.

The proof will again go via a maximal inequality which we now prove using methods taken from differentiation theory for Lebesgue integrals; namely, a discrete version of Wiener's covering lemma:

Lemma 23.2 (Maximal inequality in $\mathbb{N}^d$) *Given $h\colon \mathbb{N}^d \to \mathbb{R}$ and, for $k \geqslant 1$, abbreviate $B_k(x) := x + [0,k)^d \cap \mathbb{N}^d$. Set*

$$h_n^\star(x) := \max_{k=1,\dots,n} \frac{1}{|B_k(x)|} \sum_{z\in B_k(x)} \bigl|h(z)\bigr| \tag{23.6}$$

Then

$$\forall \lambda > 0\colon \quad \bigl|\{x \in B_n(0)\colon h_n^\star(x) > \lambda\}\bigr| \leqslant \frac{3^d}{\lambda} \sum_{z\in B_{2n}(0)} \bigl|h(z)\bigr| \tag{23.7}$$

Proof. The set $\{x \in B_n(0)\colon h_n^\star(x) > \lambda\}$ is covered by the collection of boxes

$$\Bigl\{B_k(x)\colon x \in B_n(0),\ k = 1,\dots,n,\ \sum_{x\in B_k(x)} \bigl|h(z)\bigr| > \lambda B_k(x)\Bigr\}. \tag{23.8}$$

Let $B^1, \dots, B^m$ enumerate these in decreasing order of their linear size let x_i and k_i be such that $B^i = B_{k_i}(x_i)$. Denote by

$$\widetilde{B}^i := x_i + [-k_i, 2k_i)^d \cap \mathbb{N}^d \tag{23.9}$$

be a box of three times the size centered at the same point as B^i. We construct a sequence of indices $i_1, \dots, i_q$ recursively as follows: Set $i_1 := 1$ and, assuming that $i_1, \dots, i_k$ have been defined, set $i_{k+1} := \min\{i > i_j \colon x_i \notin \widetilde{B}^{i_1} \cup \dots \cup \widetilde{B}^{i_k}\}$ if the set under the minimum is non-empty or, if otherwise, put $q := k$ and terminate. The construction implies

$$\bigl\{x \in B_n(0) \colon h_n^\star(x) > \lambda\bigr\} \subseteq \bigcup_{k=1}^m \widetilde{B}^{i_k} \tag{23.10}$$

However, the key point is that

$$B^{i_1}, \dots, B^{i_q} \text{ are disjoint} \tag{23.11}$$

which follows by noting that if we had $B^{i_k} \cap B^{i_\ell} \neq \emptyset$ for some $k < \ell$, then we would have $B^{i_\ell} \subseteq \widetilde{B}^{i_k}$ contradicting $x_{i_\ell} \notin \widetilde{B}^{i_k}$ as forced by the construction.

To prove the statement, we now note that (23.10) along with $|\widetilde{B}^{i_k}| = 3^d |B^{i_k}|$, the definition of the boxes and (23.11) gives

$$\begin{aligned} \bigl|\{x \in B_n(0) \colon h_n^\star(x) > \lambda\}\bigr| &\leqslant \sum_{k=1}^q |\widetilde{B}^{i_k}| = 3^d \sum_{k=1}^q |B^{i_k}| \\ &\leqslant \frac{3^d}{\lambda} \sum_{k=1}^q \sum_{x \in B^{i_k}} \bigl|h(z)\bigr| \leqslant \frac{3^d}{\lambda} \sum_{x \in B_{2n}(0)} \bigl|h(z)\bigr| \end{aligned} \tag{23.12}$$

where we also noted that $B^{i_1}, \dots, B^{i_q} \subseteq B_{2n}(0)$ in the last step. □

We now convert this into:

Corollary 23.3 (Maximal inequality in measure space) *Given a measure space $(\mathscr{X}, \mathcal{G}, \mu)$ with $\mu(\mathscr{X})$ finite or infinite and a family of m.p.t.'s $\{\varphi_x \colon x \in \mathbb{N}^d\}$ such that $\varphi_{x+y} = \varphi_x \circ \varphi_y$ hold for all $x, y \in \mathbb{N}^d$, we have*

$$\forall f \in L^1\, \forall \lambda > 0 \colon \quad \mu(f^\star > \lambda) \leqslant \frac{6^d}{\lambda} \|f\|_1 \tag{23.13}$$

where

$$f^\star := \sup_{n \geqslant 1} \frac{1}{|B_n(0)|} \sum_{z \in B_n(0)} |f| \circ \varphi_x \tag{23.14}$$

In particular, $f \in L^1$ implies $f^\star < \infty$ a.e.

Proof. Let $f \in L^1$ and, for $x \in \mathbb{Z}^d$ with non-negative coordinates, set $h(x) := f \circ \varphi_x$. Set

$$M_n f := \sup_{k=1,\dots,n} \frac{1}{|B_k(0)|} \sum_{z \in B_k(0)} |f| \circ \varphi_x \tag{23.15}$$

and note that then $h_n^\star(x) = M_n f \circ \varphi_x$ thanks to the assumption that $\varphi_{x+y} = \varphi_x \circ \varphi_y$. Integrating (23.7) over μ while using that φ_x is μ-preserving yields

$$\begin{aligned}
\mu(M_n f > \lambda) &= \frac{1}{|B_n(0)|} \sum_{x \in B_n(0)} \int 1_{\{M_n f > \lambda\}} \circ \varphi_x \,\mathrm{d}\mu \\
&= \frac{1}{|B_n(0)|} \int \bigl|\{x \in B_n(0) \colon M_n f > \lambda\}\bigr| \mathrm{d}\mu \\
&\leqslant \frac{3^d}{\lambda} \frac{1}{|B_n(0)|} \int \sum_{x \in B_{2n}(0)} |f| \circ \varphi_x \,\mathrm{d}\mu = \frac{6^d}{\lambda} \int |f| \mathrm{d}\mu
\end{aligned} \tag{23.16}$$

where we used that $|B_{2n}(0)| = 2^d |B_n(0)|$. Since

$$f^\star := \sup_{n \geqslant 1} M_n f \tag{23.17}$$

the Monotone Convergence Theorem then gives the claim. □

Proof of Theorem 23.1. We will proceed as in our second proof of Birkhoff's theorem, modulo caveats due to the general form of sets $\{\Lambda_n\}_{n \geqslant 1}$ and the fact that $\mu(\mathscr{X})$ is now allowed to be infinite (and L^2 may not be a subset of L^1).

Let $\varphi_1, \dots, \varphi_d$ be the m.p.t.'s as in the statement and pick f of the form

$$f = g + \sum_{i=1}^d (h_i - h_i \circ \varphi_i) \tag{23.18}$$

for some $g, h_1, \dots, h_d \in L^1 \cap L^\infty$. We claim that for any sequence $\{\Lambda_n\}_{n \geqslant 1}$ as in the statement, the limit (23.5) exists and equals g. For this it suffices to show that, for all $h \in L^1 \cap L^\infty$ and $i = 1, \dots, d$,

$$\frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} (h - h \circ \varphi_i) \circ \varphi_x \underset{n \to \infty}{\longrightarrow} 0 \tag{23.19}$$

This follows from the fact that, by writing the sum over the i-th coordinate first and integrating the gradients, we have

$$\biggl| \sum_{x \in \Lambda_n} (h - h \circ \varphi_i) \circ \varphi_x \biggr| \leqslant \|h\|_\infty |\partial \Lambda_n| \tag{23.20}$$

and so we get (23.19) from $|\partial \Lambda_n| / |\Lambda_n| \to 0$ as implied by (23.4).

Next we claim that

$$\biggl\{ f \in L^1 \colon \lim_{n \to \infty} \frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} f \circ \varphi_x \text{ exists a.e.} \biggr\} \text{ is closed in } L^1 \tag{23.21}$$

Let $\{f_k\}_{k \geqslant 1}$ be a sequence in this set such that $f_k \to f$ in L^1. Abbreviating $A_n f := \frac{1}{|\Lambda_n|} \sum_{x \in \Lambda_n} f \circ \varphi_x$, on $\{|f_k - f|^\star \leqslant \epsilon\}$ the containment $\Lambda_n \subseteq B_n(0)$ gives

$$\forall n \geqslant 1 \colon \quad A_n f_k - \epsilon \leqslant A_n f \leqslant A_n f_k + \epsilon \tag{23.22}$$

and so

$$\bar{f}_k - \epsilon \leqslant \liminf_{n\to\infty} A_n f \leqslant \limsup_{n\to\infty} A_n f \leqslant \bar{f}_k + \epsilon \quad \text{a.e.} \tag{23.23}$$

where $\bar{f}_k$ denotes the a.e. value of $\lim_{n\to\infty} A_n f_k$. Noting that the Maximal Inequality (23.13) implies $\bar{f}_k$ is finite a.e., we may subtract the left-hand side from the right-hand side to get

$$\begin{aligned}\mu\Bigl(\limsup_{n\to\infty} A_n f - \liminf_{n\to\infty} A_n f > 2\epsilon\Bigr)& \\ \leqslant \mu\bigl(\,|f_k - f|^\star > \epsilon\bigr) \leqslant \frac{6^d}{\epsilon}\|f_k - f\|_1&\end{aligned} \tag{23.24}$$

by one additional application of (23.13). Taking $k \to \infty$ followed by $\epsilon \downarrow 0$ we conclude that $\lim_{n\to\infty} A_n$ exists and is finite a.e. as well, proving (23.21).

To prove the convergence part of the claim, we again observe that, by the argument underlying the proof of the Mean Ergodic Theorem,

$$\mathscr{L} := \left\{ g + \sum_{i=1}^d (h_i - h_i \circ \varphi_i) \colon g, h_1, \dots, h_d \in L^1 \cap L^\infty \wedge \forall i = 1, \dots, d\colon\ g = g \circ \varphi_i \right\} \tag{23.25}$$

is dense in L^2. Let $f \in L^1 \cap L^\infty$. The Maximal Inequality then ensures that

$$E_k := \bigl\{\limsup_{n\to\infty} |A_n f| > 1/k\bigr\} \tag{23.26}$$

is an invariant set with $\mu(E_k) < \infty$ for each k. Assuming k is such that $\mu(E_k) > 0$, the function $f_k := f 1_{E_k}$ is then in L^2 and so, given $\epsilon > 0$, there exist $f' \in \mathscr{L}$ such that

$$\|f_k - f'\|_2 < \epsilon[1 + \mu(E_k)]^{-1/2} \tag{23.27}$$

But $f' = g + \sum_{i=1}^d (h_i - h_i \circ \varphi_i)$ and so, denoting $\tilde{f}' := f 1_{E_k}$, we then have $\tilde{f}' = \tilde{g} + \sum_{i=1}^d (\tilde{h}_i - \tilde{h}_i \circ \varphi_i)$ with $\tilde{g} := g 1_{E_k}$ and $\tilde{h}_i := h_i 1_{E_k}$; i.e., $\tilde{f}' \in \mathscr{L}$ as well. Moreover, $\|f_k - \tilde{f}'\|_2 < \epsilon[1 + \mu(E_k)]^{-1/2}$ holds by restricting the integration to E_k. As $f_k - \tilde{f}'$ is supported in the finite measure set E_k, the Cauchy-Schwarz inequality gives

$$\|f_k - \tilde{f}'\|_1 \leqslant \mu(E_k)^{1/2} \|f_k - f'\|_2 < \epsilon \frac{\mu(E_k)^{1/2}}{[1 + \mu(E_k)]^{1/2}} < \epsilon \tag{23.28}$$

As this holds for all $\epsilon > 0$, we conclude that f_k lies in L^1-closure of $\mathscr{L}$. Noting that $A_n f_k = 1_{E_k} A_n f$, from $\mathscr{L}$ being a subset of the set in (23.21) we then get that $\lim_{n\to\infty} A_n f$ exists a.e. on E_k, for each $k \geqslant 1$.

In light of

$$\mathscr{X} \smallsetminus \bigcup_{k\geqslant 1} E_k = \{\limsup_{n\to\infty} |A_n f| = 0\} \tag{23.29}$$

we have that $\lim_{n\to\infty} A_n f$ exists a.e., and the set in (23.21) contains, all $f \in L^1 \cap L^\infty$. Since $L^1 \cap L^\infty$ is dense in L^1 by Dominate Convergence, the a.e. convergence takes places for all $f \in L^1$, as we desired to show. The proof that the convergence is in L^1 when $\mu(\mathscr{X}) < \infty$ is as in Theorem 22.1 and so we omit it. □

23.2 Dominated Ergodic Theorem.

Our study of ergodic theorems continues by analysis of integrability of the maximum function. We start by noting the standard fact we already saw for martingles:

Lemma 23.4 *Let $f \mapsto f^\star$ be a map on measurable functions that is monotone (i.e., $f \leqslant g$ implies $f^\star \leqslant g^\star$), positive-homogeneous and subadditive (i.e., $(af+bg)^\star \leqslant |a|f^\star + |b|g^\star$) and acts as identity on positive constants (i.e., $1^\star = 1$). Suppose that*

$$\forall \lambda > 0\, \forall f \in L^1\colon \quad \mu(f^\star > \lambda) \leqslant \frac{C}{\lambda}\|f\|_1 \tag{23.30}$$

holds. Then for each $p \in (1,\infty)$,

$$\forall f \in L^p\colon \quad \|f^\star\|_p \leqslant 2C^{1/p}\Bigl(\frac{p}{p-1}\Bigr)^{1/p}\|f\|_p \tag{23.31}$$

Proof. Assume without loss of generality that $f \geqslant 0$. For $\lambda > 0$ set $g := f1_{\{f \geqslant \lambda/2\}}$. Then $g \leqslant f + \lambda/2$ implies $g^\star \leqslant f^\star + \lambda/2$. Hence

$$\mu(f^\star > \lambda) \leqslant \mu(g^\star > \lambda/2) \leqslant \frac{2C}{\lambda}\int g\mathrm{d}\mu = \frac{2C}{\lambda}\int_{\{f>\lambda/2\}} f\mathrm{d}\mu \tag{23.32}$$

This gives

$$\begin{aligned}\int (f^\star)^p\mathrm{d}\mu &= \int_0^\infty p\lambda^{p-1}\mu(f^\star > \lambda)\mathrm{d}\lambda \\ &\leqslant 2Cp\int_0^\infty \lambda^{p-1}\Bigl(\int f1_{\{f>\lambda/2\}}\mathrm{d}\mu\Bigr)\mathrm{d}\lambda \\ &= 2Cp\int\Bigl(\int_0^\infty \lambda^{p-2}1_{\{f>\lambda/2\}}\mathrm{d}\lambda\Bigr)f\mathrm{d}\mu = \frac{2Cp}{p-1}\int (2f)^{p-1}f\mathrm{d}\mu\end{aligned} \tag{23.33}$$

where we used Tonelli's theorem to swap the order of integration. Now collect the constants to write the right hand side as $2^pC\frac{p}{p-1}\int f^p\mathrm{d}\mu$. □

A natural question is how bad is integrability of $f^\star$ compared to that of f. The following provides a sufficient condition:

Theorem 23.5 (Dominated Ergodic Theorem, Wiener 1939) *Assuming the setting of Theorem 23.1 with $\mu(\mathscr{X}) < \infty$, let $f^\star$ be the maximal function* (23.14). *Denote*

$$L\log L := \Bigl\{f \in L^1\colon \int |f|\log_+|f|\mathrm{d}\mu < \infty\Bigr\} \tag{23.34}$$

(This is sometimes called the Zygmund class.) Then

$$f \in L\log L \quad \Rightarrow \quad f^\star \in L^1 \tag{23.35}$$

and so $f \in L\log L$ implies that the ergodic averages admit an integrable dominating function.

Proof. Using that $\mu(\mathscr{X}) < \infty$ we write

$$\begin{aligned}\int f^\star \mathrm{d}\mu &= \int_0^\infty \mu(f^\star > \lambda)\mathrm{d}\lambda \\ &\leqslant 2\mu(\mathscr{X}) + \int_2^\infty \mu(f^\star > \lambda)\mathrm{d}\lambda\end{aligned} \tag{23.36}$$

For the integral on the right we invoke the inequality (23.21) to get

$$\begin{aligned}\int_2^\infty \mu(f^\star > \lambda)\mathrm{d}\lambda &\leqslant 2C\int_2^\infty \biggl(\int |f| 1_{\{|f|>\lambda/2\}}\mathrm{d}\mu\biggr)\mathrm{d}\lambda \\ &= 2C\int \biggl(\int_2^\infty 1_{\{|f|>\lambda/2\}}\mathrm{d}\lambda\biggr)|f|\mathrm{d}\mu = 2C\int |f|\log|f|\,\mathrm{d}\mu\end{aligned} \tag{23.37}$$

The integral on the right is bounded by that with log replaced by its positive part. □

Since the convergence of ergodic averages in L^1 can be proved directly, the main use of the above is a sufficient condition for having an integrable function that dominates the ergodic averages. Note that the statement is generally false when $\mu(\mathscr{X}) = \infty$ for if we had $f^\star \in L^1$, then the convergence in the Spatial Ergodic Theorem would take place in L^1 which is generally false in infinite measure spaces. As shown by K. Petersen in 1979, for continuous analogues of ergodic averages (i.e., expressions of the type $\frac{1}{T}\int_0^T f\circ\varphi_t \mathrm{d}t$) the Maximal Inequality is in fact an equality and the computation above can be reversed to show that $f^\star \in L^1$ implies $f \in L\log L$.

An interesting twist is that all one actually needs is that the maximal function is finite almost surely. This is a consequence of:

Theorem 23.6 (S. Banach, 1925) *Let $p \in [1,\infty]$ and let $(\mathscr{X},\mathcal{G},\mu)$ be a measure space. Assume $\{T_n\}_{n\geqslant 1}$ are bounded linear operators on L^p such that*

$$\forall f \in L^p\colon \quad T^\star f := \sup_{n\geqslant 1} Tf < \infty \quad \mu\text{-a.e.} \tag{23.38}$$

Then

$$\Bigl\{f \in L^p\colon \lim_{n\to\infty} T_n f \text{ exists } \mu\text{-a.e.}\Bigr\} \text{ is closed in } L^p \tag{23.39}$$

The proof uses a reasoning similar to the proof of the Uniform Boundedness Principle to show that a maximal inequality of sorts holds once $T^\star f$ is finite almost surely for all $f \in L^p$. We leave the details to an exercise.

23.3 Some other ergodic theorems.

We now state two additional convergence theorems. The first one is concerns with compositions of m.p.t.'s that are themselves chosen at random:

Theorem 23.7 (Random Ergodic Theorem, S. Ulam and J. von Neumann 1945) *Given a measure space $(\mathscr{X},\mathcal{G},\mu)$, a measurable space (S,Σ) and a map $\varphi\colon \mathscr{X}\times S \to \mathscr{X}$ satisfying*

(1) *for each $y \in S$, the map $x \mapsto \varphi(x,y)$ is a m.p.t. on $(\mathscr{X},\mathcal{G},\mu)$,*
(2) *$(x,y) \mapsto \varphi(x,y)$ is $\mathcal{G}\otimes\Sigma/\mathcal{G}$-measurable,*

abbreviate the n-fold composition with second coordinates $y_0, \dots, y_{n-1}$ as

$$\varphi_n(x, y_0, \dots, y_{n-1}) := \varphi(\cdot, y_{n-1}) \circ \cdots \circ \varphi(\cdot, y_0)(x) \tag{23.40}$$

For any sequence $\{Y_k\}_{k \geqslant 0}$ of i.i.d. S-valued random variables on a probability space $(\Omega, \mathcal{F}, P)$ and any $f \in L^1(\mathscr{X}, \mathcal{G}, \mu)$,

$$\bar{f} := \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f \circ \varphi_k(\cdot, Y_0, \dots, Y_{k-1}) \text{ exists a.e., } P\text{-a.s.} \tag{23.41}$$

Moreover, we have $\bar{f} \in L^1(\mathscr{X} \times \Omega, \mathcal{G} \times \mathcal{F}, \mu \otimes P)$.

We remark that, as far as convergence is concerned, it suffices that $\{Y_k\}_{k \geqslant 0}$ are stationary (with respect to the left shift). The proof is a good exercise on the concept of *skew product*. We leave the proof to a homework exercise.

Another theorem concerns a subtlety of potentially dealing with uncountably many null sets when a continuum-valued parameter is involved:

Theorem 23.8 (N. Wiener and A. Wintner, 1941) *Let φ be a m.p.t. on $(\mathscr{X}, \mathcal{G}, \mu)$ with $\mu(\mathscr{X}) < \infty$. For all $f \in L^1(\mu)$ there exists a μ-null set $\mathcal{N} \in \mathcal{G}$ such that*

$$\forall x \in \mathscr{X} \smallsetminus \mathcal{N} \; \forall \theta \in \mathbb{R}: \quad \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathrm{e}^{2\pi \mathrm{i} \theta k} f \circ \varphi^k(x) \text{ exists} \tag{23.42}$$

A key point is that the set $\mathcal{N}$ can be chosen uniformly for all θ. It is instructive to first prove convergence in the mean. This goes by constructing a set of functions which is dense in L^1 and for which convergence is verified explicitly. Then we observe that the maximal function $f^\star$ dominates the ergodic averages uniformly in θ. We leave the details to a homework exercise.