CHAPTER 11

MIXING PROPERTIES

We now want to consider two stronger properties than ergodicity. These are weak mixing and strong mixing which are important from the statistical point of view, as we shall see in the next chapter.

11.1 Weak mixing

DEFINITION. Let $T : X \to X$ be a measure preserving map on a probability space $(X, \mathcal{B}, \mu)$; then we call T *weak-mixing* if for any $A, B \in \mathcal{B}$ we have that

$$\frac{1}{N}\sum_{n=0}^{N-1} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \to 0$$

as $N \to +\infty$.

We have the following equivalent characterization.

LEMMA 11.1. *The following are equivalent.*

(i) T *is weak-mixing;*

(ii) *for* $f, g \in L^2(X, \mathcal{B}, \mu)$ *we have that*

$$\frac{1}{N}\sum_{n=0}^{N-1} |\int fT^n g d\mu - \int f d\mu \int g d\mu| \to 0$$

as $N \to +\infty$.

PROOF. For "(ii) implies (i)" we need only make the choices $f = \chi_A$ and $g = \chi_B$. For "(i) implies (ii)" we can use an argument of approximation by step functions (finite linear combinations of characteristic functions).

■

The following lemma shows that weak-mixing is a stronger property than ergodicity.

LEMMA 11.2. *If a transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$ is weak-mixing then it is necessarily ergodic.*

PROOF. If T is weak-mixing then by definition we have that for any $A, B \in \mathcal{B}$

$$\frac{1}{N}\sum_{n=0}^{N-1} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \to 0$$

as $N \to +\infty$. By the triangle inequality we have that

$$\begin{aligned} &|\frac{1}{N}\sum_{n=0}^{N-1} \mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \\ &\leq \frac{1}{N}\sum_{n=0}^{N-1} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \\ &\to 0. \end{aligned} \tag{11.1}$$

If we assume (for a contradiction) that T were not ergodic then there would exist a T-invariant set $E \in \mathcal{B}$ with $T^{-1}E = E$ with $0 < \mu(E) < 1$. If we take $A = E$ and $B = X - E$ in (11.1) then since $\mu(T^{-n}E \cap (X - E)) = \mu(E \cap (X - E))$, for all $n \geq 0$, we deduce that $\mu(E) \cdot \mu(X - E) = 0$ giving the required contradiction. Thus T is ergodic.

■

The converse is not true: there exist examples of transformations which are ergodic but not weak-mixing, as the following simple example shows.

EXAMPLE (ERGODIC, NOT WEAK-MIXING). Let $X = \mathbb{R}/\mathbb{Z}$, let $\mathcal{B}$ be the Borel sigma-algebra, and let μ be the Haar-Lebesgue measure. For any irrational number $a \in \mathbb{R}$ the transformation $T : X \to X$ defined by $T(x) = x + a \pmod 1$ is known to be ergodic. We can see that it is not weak-mixing by choosing $A = B = [0, \frac{1}{2}]$, and then $\mu(A)\mu(B) = \frac{1}{4}$. Since the sequence $na + \mathbb{Z}$ is uniformly distributed we know that the proportion of the terms in the sub-sequence n_i for which $n_i a \in [0, \frac{1}{100}] \pmod 1$ is $\frac{1}{100}$. For these terms we have that $\mu(T^{n_i}A \cap B) - \mu(A)\mu(B) \geq \frac{49}{100} - \frac{1}{4} = \frac{24}{100}$ which means that

$$\liminf_{N \to +\infty} \frac{1}{N}\sum_{n=0}^{N-1} |\mu(T^{n_i}A \cap B) - \mu(A)\mu(B)| \geq \frac{1}{100}.\frac{24}{100} > 0.$$

11.2 A density one convergence characterization of weak mixing

Using the previous lemma on sequences there is also a characterization for weak mixing which is closer to that of strong mixing. We say that a sequence $\{n_i\}_{i \in \mathbb{N}}$ of natural numbers has density one if

$$\lim_{n \to +\infty} \frac{1}{n}\mathrm{Card}\{n_i \in [0, 1, \ldots, n-1] : i \in \mathbb{N}\} = 1.$$

PROPOSITION 11.3. *The transformation T is weak-mixing if there exists a sequence $\{n_i\}_{i\in\mathbb{N}}$ of density one such that $\mu(T^{-n_i}A \cap B) \to \mu(A)\mu(B)$ as $i \to +\infty$.*

The proof requires only the following simple lemma on sequences.

LEMMA 11.4. *The following are equivalent for a bounded sequence of real positive numbers $\{a_n\}$.*

(1) $\frac{1}{n}\sum_{k=0}^{n-1} a_k \to 0$ *as* $n \to +\infty$; *and*
(2) $\lim_{k\to+\infty} a_{n_k} = 0$ *for some sub-sequence* $\{n_k\} \subset \mathbb{N}$ *of density one (i.e.* $\lim_{n\to+\infty} \frac{1}{n}\mathrm{Card}\{n_i \in [0, 1, \ldots, n-1]\colon\ i \in \mathbb{N}\} = 1$*).*

PROOF. (2) $\Longrightarrow$ (1): Let J be a sequence of density one. Given $\epsilon > 0$ we can choose N such that for $n \geq N$ we have that

$$Card\{n_i \in [0, 1, \ldots, n-1]\} \geq n(1-\epsilon)$$

and for $n_i \geq N$ we have $a_{n_i} \leq \epsilon$. In particular,

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1} a_k &= \frac{1}{n}\left(\sum_{k=0}^{N} a_k + \sum_{\substack{k\in\{n_i\}\\ N<k\leq n-1}} a_k + \sum_{\substack{k\notin\{n_i\}\\ N<k\leq n-1}} a_k\right)\\ &\leq \frac{1}{n}\sum_{k=0}^{N} a_k + \epsilon\left((1-\epsilon) + \sup\{a_i\}\right).\end{aligned}$$

We can choose n sufficiently large that $\frac{1}{n}\sum_{k=0}^{N} a_k < \epsilon$. Since $\epsilon > 0$ can be chosen arbitrarily small we have that $\frac{1}{n}\sum_{k=0}^{n-1} a_k \to 0$.

(1) $\Longrightarrow$ (2): Assume that $\frac{1}{n}\sum_{k=0}^{n-1} a_k \to 0$. For each $m \geq 1$ we define

$$J_m = \{n \in \mathbb{N} :\ a_n \leq \frac{1}{m}\}$$

and observe that this has density one, since

$$\frac{1}{m}\left(\frac{1}{n}\sum_{k=0}^{n-1} \chi_{\mathbb{N}-J_m}(k)\right) \leq \frac{1}{n}\sum_{k=0}^{n-1} a_k \to 0$$

as $n \to +\infty$. Thus for each $m \geq 1$ we can choose n_m such that

$$\frac{1}{n}\left(\sum_{k=0}^{n-1} \chi_{\mathbb{N}-J_m}(k)\right) \leq \frac{1}{m}$$

for $n \geq n_m$. We then define

$$J = \cup_{k=1}^{\infty} J_k \cap [n_k, \ldots, n_{k+1}]$$

and it is easy to see that J has density one and $\lim_{n\to\infty:n\in J} a_n = 0$.

∎

The following corollary is quite useful.

COROLLARY 11.4.1. *$\frac{1}{n}\sum_{k=0}^{n-1} a_k \to 0$ as $n \to +\infty$ if and only if $\frac{1}{n}\sum_{k=0}^{n-1} a_k^2 \to 0$ as $n \to +\infty$.*

11.3 A generalization of the von Neumann ergodic theorem

We want to present the following interesting generalization of the Von Neumann ergodic theorem (Theorem 10.1) for weak-mixing transformations. It will only be used in chapter 16 and is not required for the rest of this chapter.

THEOREM 11.5. *Assume that $f_1, \dots, f_k \in L^\infty(X, \mathcal{B}, \mu)$. If $T : X \to X$ is weak-mixing then*

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \to \int f_1 d\mu \int f_2 d\mu \dots \int f_k d\mu$$

(in the L^2 topology) as $N \to +\infty$.

PROOF. The proof is by induction. When $k = 1$, this is precisely the Von Neumann ergodic theorem (Theorem 10.1).

Assume that the result has been established for $k-1$ functions. We may assume without loss of generality that $\int f_k d\mu = 0$ (otherwise we need only replace f_k by $f_k - \int f_k d\mu$). Thus it suffices to show that

$$\int |\frac{1}{N}\sum_{n=1}^{N} f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x)|^2 d\mu(x) \to 0 \text{ as } N \to +\infty.$$

For any $1 \le m \le N$ we can now bound

$$\begin{aligned}
&\int |\frac{1}{N}\sum_{n=1}^{N} f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x)|^2 d\mu(x) \\
&\le \int \left| \frac{1}{N}\sum_{n=1}^{N} \left(\frac{1}{m}\sum_{j=0}^{m-1} f_1(T^{n+j} x) f_2(T^{2(n+j)} x) \dots f_k(T^{k(n+j)} x) \right) \right|^2 d\mu(x) \\
&\qquad + \left(\frac{2m}{N} + \frac{m^2}{N^2} \right) \left(\max_{1 \le i \le k} |f_i|_\infty \right)^2 \\
&\le \frac{1}{N}\sum_{n=1}^{N} \left(\int |\frac{1}{m}\sum_{j=0}^{m-1} f_1(T^{n+j} x) f_2(T^{2(n+j)} x) \dots f_k(T^{k(n+j)} x)|^2 d\mu(x) \right) \\
&\qquad + \left(\frac{2m}{N} + \frac{m^2}{N^2} \right) \left(\max_{1 \le i \le k} |f_i|_\infty \right)^2 .
\end{aligned} \tag{11.2}$$

(The first inequality comes from the observation that

$$\frac{1}{N}\sum_{n=1}^{N} a_n \leq \frac{1}{N}\sum_{n=1}^{N-(m-1)}\left(\frac{1}{m}\sum_{j=0}^{m-1} a_{n+j}\right) + \frac{m}{N}\left(\max_{1\leq i\leq m}|a_i| + \max_{N-m\leq i\leq N}|a_i|\right)$$

for any real numbers $a_1, \ldots, a_N$. The second inequality comes from the observation that $\left(\frac{1}{N}\sum_{n=1}^{N} b_n\right)^2 \leq \frac{1}{N}\sum_{n=1}^{N}|b_n|^2$ for any real numbers $b_1, \ldots, b_N$.) We next observe that

$$\begin{aligned}
&\int |\sum_{j=0}^{m-1} f_1(T^{n+j}x)f_2(T^{2(n+j)}x)\ldots f_k(T^{k(n+j)}x)|^2 d\mu(x) \\
&= \sum_{i=0}^{m-1}\sum_{j=0}^{m-1}\int\left(\prod_{l=1}^{k} f_l(T^{l(n+i)}x)\right)\left(\prod_{l=1}^{k} f_l(T^{l(n+j)}x)\right)d\mu(x) \qquad (11.3)\\
&= \sum_{i=0}^{m-1}\sum_{j=0}^{m-1}\int\left(\prod_{l=1}^{k}\left(f_l.f_l\circ T^{l(j-i)}\right)(T^{l(n+i)}x)\right)d\mu(x).
\end{aligned}$$

By the inductive hypothesis we know that for each $0 \leq i, j \leq m-1$,

$$\frac{1}{N}\sum_{n=1}^{N}\prod_{l=2}^{k}\left(f_l\cdot f_l\circ T^{l(j-i)}\right)(T^{l(n+i)}x) \to \prod_{l=2}^{k}\int f_l\cdot f_l\circ T^{l(j-i)}d\mu$$

as $N \to +\infty$ (in the L^2 topology) and so

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\int\prod_{l=1}^{k}\left(f_l.f_l\circ T^{l(j-i)}\right)(T^{l(n+i)}x)d\mu(x)\\
&= \frac{1}{N}\sum_{n=1}^{N}\int\left(f_1\cdot f_1\circ T^{(j-i)}\right)(x)\left(\prod_{l=2}^{k}\left(f_l\cdot f_l\circ T^{l(j-i)}\right)(T^{l(n+i)}x)\right)d\mu(x)\\
&\to \prod_{l=1}^{k}\int f_l\cdot f_l\circ T^{l(j-i)}d\mu.
\end{aligned} \qquad (11.4)$$

Comparing (11.2), (11.3) and (11.4) we see that

$$\begin{aligned}
&\limsup_{N\to+\infty}\int|\frac{1}{N}\sum_{n=1}^{N} f_1(T^n x)f_2(T^{2n}x)\ldots f_k(T^{kn}x)|^2 d\mu(x)\\
&\leq \frac{1}{m^2}\sum_{i=0}^{m-1}\sum_{j=0}^{m-1}\left(\prod_{l=1}^{k}\int f_l\cdot f_l\circ T^{l(j-i)}d\mu(x)\right).
\end{aligned} \qquad (11.5)$$

Finally, since T is weak-mixing we know that $\int f_l\cdot f_l\circ T^{lr_n}d\mu(x) \to \int f_l d\mu(x) = 0$ where $r_n \to +\infty$ through a set of density one. Thus for sufficiently large m the expression in (11.4) can be made arbitrarily small.

■

11.4 The spectral viewpoint

Consider a measure preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$. Consider the Hilbert space $H = L^2(X, \mathcal{B}, \mu)$ of square integrable functions with the inner product $\langle f, g\rangle = \int f\bar{g}d\mu$. We can associate to T an operator $U_T : H \to H$ defined by $(U_T f)(x) = f(Tx)$ whenever $f \in H$. It is easy to see the following.

LEMMA 11.6. *$U_T : H \to H$ is an isometry. i.e. $||U_T f|| = ||f||$.*

We recall a few elementary observations about operators $U : H \to H$ on a Hilbert space H. We call a linear operator U an *isometry* if for every pair of vectors $x, y \in H$ we have that $\langle Ux, Uy\rangle = \langle x, y\rangle$. We shall only be interested in isometries.

An eigenvalue for $U : H \to H$ is a complex number $\alpha \in \mathbb{C}$ for which there exists a (non-zero) vector $x \in H$ (called the eigenvector) such that $Ux = \alpha x$.

LEMMA 11.7. *Eigenvalues of isometries must be complex numbers of modulus unity.*

PROOF. Clearly, if $Ux = \alpha x$ then $\langle Ux, Ux\rangle = \alpha\bar{\alpha}\langle x, x\rangle$. But since U is an isometry we have that $\langle Ux, Uy\rangle = \langle x, y\rangle$ and so $\langle x, x\rangle = |\alpha|^2\langle x, x\rangle$ i.e. $|\alpha| = 1$, as claimed.

An important aspect of the spectrum of the operator is the variety (or lack of it) of eigenvectors. Two extreme cases are the following.

DEFINITION. The operator $U : H \to H$ has *continuous spectrum* if there are no eigenvectors. The operator $U : H \to H$ has *pure point spectrum* if H is the closure of the linear span of the eigenvectors.

REMARK. Between these two extreme cases we have the possibility of having *mixed spectrum*. We can let $V \subset H$ denote the subspace spanned by the eigenvectors. We let $V^\perp \subset H$ denote the orthogonal subspace to V. ¿From the definitions we see that $U : V \to V$ then has pure point spectrum and $U : V^\perp \to V^\perp$ has continuous spectrum.

We now recall one of the basic theorems in spectral theory.

DEFINITION. A sequence $r_n \in \mathbb{C}$, $n \in \mathbb{Z}$, is called *positive definite* if for each $N \geq 1$ and each sequence $a_0, \ldots, a_N$ we always have that

$$\sum_{n,m=0}^{N} r_{n-m} a_n \bar{a}_m \geq 0.$$

BOCHNER-HERGLOTZ SPECTRAL THEOREM. *If r_n, $n \in \mathbb{Z}$, is positive definite then there is a unique finite Borel measure μ on $\mathbb{R}/\mathbb{Z}$ such that $r_n = \int_0^1 e^{2\pi int} d\mu(t)$.*

(The proof can be found in the appendix to [1].)

APPLICATION TO ISOMETRIES. The Bochner-Herglotz theorem is particularly well suited to isometries $U : H \to H$. Fix $x \in H$ and then set $r_n = \langle U^n x, x\rangle$ and $r_{-n} = \langle x, U^n x\rangle$ and observe that

$$\sum_{n,m=0}^{N} r_{n-m} a_n \bar{a}_m = \langle \sum_{n=0}^{N} a_n U^n x, \sum_{n=0}^{N} a_n U^n x\rangle \geq 0.$$

The measure μ on the unit circle $\mathbb{R}/\mathbb{Z}$ is called the *spectral measure.*

The choice of point $x \in H$ affects the resulting spectral measure μ. If $x \in V$ (i.e. x is in the closure of the span of the eigenvectors) then the associated measure is singular with respect to the Haar-Lebesgue measure on the circle $\mathbb{R}/\mathbb{Z}$. If $x \in V^{\perp}$ then the associated measure is absolutely continuous with respect to the Haar-Lebesgue measure on the circle $\mathbb{R}/\mathbb{Z}$.

EXAMPLE. Consider an irrational rotation $T : X \to X$ on the unit circle $X = \mathbb{R}/\mathbb{Z}$ defined by $T(x + \mathbb{Z}) = (x + a + \mathbb{Z})$. The functions $e_n(x) = e^{2\pi inx}$ are eigenfunctions for the operator $U : L^2(X, \mathcal{B}, \mu) \to L^2(X, \mathcal{B}, \mu)$ since

$$Ue_n(x) = e^{2\pi in(x+a)} = e^{2\pi ina} e_n(x).$$

Since the family $e_n, n \in \mathbb{Z}$, spans the space the transformation T has pure point spectrum.

In applying this to ergodic theory, we consider a measure preserving transformation $T : X \to X$ on the Hilbert space $H = L^2(X, \mathcal{B}, \mu)$ with inner product $< f, g >= \int f\bar{g}d\mu$.

REMARK. The following lemma is also a standard result from spectral theory (although we won't require it).

RIEMANN-LEBESGUE LEMMA. *If the spectral measure μ on $\mathbb{R}/\mathbb{Z}$ for the operator $U : H \to H$ (and a point $x \in H$) is absolutely continuous then $< U^n x, x >\to 0$ as $n \to +\infty$.*

11.5 Spectral characterization of weak mixing

The Hilbert space $H = L^2(X, \mathcal{B}, \mu)$ has the obvious one-dimensional subspace consisting of constant functions and denoted by $\mathbb{C}$. We let $\mathbb{C}^\perp$ denote the orthonormal (co-dimension one) subspace. For a measure preserving transformation $T : X \to X$ the associated isometry $U : H \to H$ preserves both $\mathbb{C}$ and $\mathbb{C}^\perp$.

PROPOSITION 11.8. *Let $T : X \to X$ be a measure preserving transformation on the probability space $(X, \mathcal{B}, \mu)$. The following conditions are equivalent:*

(1) *for the map $U : \mathbb{C}^\perp \to \mathbb{C}^\perp$ has continuous spectrum;*
(2) *T is weak-mixing;*
(3) *the measure preserving transformation $T \times T : X \times X \to X \times X$ (on $X \times X$ with the product measure $\mu \times \mu$) defined by $(T \times T)(x, y) = (Tx, Ty)$ is weak-mixing (and thus ergodic).*

PROOF.

(1) $\Longrightarrow$ (2): Assume that $U : \mathbb{C}^\perp \to \mathbb{C}^\perp$ has continuous spectrum. $\bar{\mu}$ denotes the spectral meausre. Choose any vector $f \in \mathbb{C}^\perp$; then we estimate that

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=0}^{N-1} |\int f \circ T^n \bar{f} d\mu|^2 \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \left(\int f \circ T^n \bar{f} d\mu\right)\left(\int \bar{f} \circ T^n f d\mu\right) \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \left(\int_0^1 e^{2\pi i n t} d\bar{\mu}(t)\right)\left(\int_0^1 e^{-2\pi i n s} d\bar{\mu}(s)\right) \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \int_0^1 \int_0^1 e^{2\pi i n(t-s)} d(\bar{\mu} \times \bar{\mu})(t, s) \\
&= \int_0^1 \int_0^1 \frac{1}{N}\left(\frac{e^{2\pi i N(t-s)} - 1}{e^{2\pi i(t-s)} - 1}\right) d(\bar{\mu} \times \bar{\mu})(t, s).
\end{aligned}
$$

Observe that since μ has continuous spectrum the product measure $\bar{\mu} \times \bar{\mu}$ gives the diagonal $\{(t, t) : t \in \mathbb{R}/\mathbb{Z}\}$ measure zero. In particular, the last integrand is finite almost everywhere. Since

$$
\frac{1}{N}\frac{e^{2\pi i N(t-s)} - 1}{e^{2\pi i(t-s)} - 1} \to 0
$$

for almost all (s,t) and it is dominated by the constant function 1, we see by the Lebesgue dominated convergence theorem that

$$\frac{1}{N}\sum_{n=0}^{N-1}|\int f\circ T^n\bar{f}d\mu|^2 \le \int_0^1\int_0^1\left(\frac{1}{N}\frac{e^{2\pi iN(t-s)}-1}{e^{2\pi i(t-s)}-1}\right)d(\bar{\mu}\times\bar{\mu})(t,s)\to 0$$

as $N\to+\infty$. Finally, we observe that

$$\frac{1}{N}\sum_{n=0}^{N-1}|\int f\circ T^n\bar{f}d\mu|^2\to 0$$

implies

$$\frac{1}{N}\sum_{n=0}^{N-1}|\int f\circ T^n\bar{f}d\mu|\to 0$$

(by Corollary 11.4.1).

(2) $\implies$ (3): Consider sets $E=\sup_i(A_i\times B_i), F=\sup_j(C_j\times D_j)$ (for finite disjoint unions of product sets A_i, B_i, C_j and D_j). Since T is weak-mixing we know that

$$\mu\left(T^{-n_k}A_i\cap C_j\right)\to\mu(A_i)\mu(C_j) \tag{11.6}$$

and

$$\mu\left(T^{-n_k}B_i\cap D_j\right)\to\mu(B_i)\mu(D_j) \tag{11.7}$$

for sequences $n_k\to+\infty$ of density one (and without loss of generality we can assume that we have the same sequence in each case).

To show $T\times T$ is weak-mixing we want to show that

$$(\mu\times\mu)\left((T\times T)^{-n_i}E\cap F\right)\to(\mu\times\mu)(E)(\mu\times\mu)(F)$$

for a sequence $n_k\to+\infty$ of unit density. This follows from (11.5) and (11.6) since

$$\begin{aligned}&(\mu\times\mu)\left((T\times T)^{-n_i}(A_i\times B_i)\cap(C_j\times D_j)\right)\\&=\mu\left(T^{-n_i}A_i\cap C_j\right)\cdot\mu\left(T^{-n_i}B_i\cap D_j\right)\\&\to\mu(A_i)\mu(C_j)\mu(B_i)\mu(D_j)\\&=\mu\left(T^{-n_i}A_i\cap C_j\right)\cdot\mu\left(T^{-n_i}B_i\cap D_j\right)\end{aligned}$$

for a sequence $n_k\to+\infty$ of unit density.

(3) $\implies$ (1): Assume for a contradiction that there is a *non-constant* eigenfunction $f\in H$ for $T:X\to X$, i.e. $Uf=\alpha f$. We can then define a function $F:X\times X\to\mathbb{C}$ by $F(x,y)=f(x)\bar{f}(y)$. Observe that

$$F(Tx,Ty)=f(Tx)\bar{f}(Ty)=\alpha\bar{\alpha}f(x)\bar{f}(y)=F(x,y).$$

But then $F(x,y)$ is a $(T\times T)$-invariant function which is non-constant. This contradicts $T\times T$ being ergodic (and therefore being weak-mixing).

11.6 Strong mixing

We now turn to another notion of "mixing".

Definition. Let $T : X \to X$ be a measure preserving transformation on a probability space $(X, \mathcal{B}, \mu)$; then we call T *strong mixing* if for any $A, B \in \mathcal{B}$ we have that

$$\mu(T^{-n}A \cap B) \to \mu(A)\mu(B)$$

as $n \to +\infty$. We have the following equivalent characterization.

Lemma 11.9. *The following are equivalent:*

(i) *T is strong-mixing;*
(ii) *for $f, g \in L^2(X, \mathcal{B}, \mu)$, $\int f \circ T^n g d\mu \to \int f d\mu \int g d\mu$ as $n \to +\infty$.*

Proof. For "(ii) $\Longrightarrow$ (i)" we need only make the choices $f = \chi_A$ and $g = \chi_B$. For "(i) $\Longrightarrow$ (ii)" we can use an argument of approximation by step functions (finite linear combinations of characteristic functions).

■

The following lemma states the obvious fact that strong mixing is a stronger property than weak mixing.

Lemma 11.10. *If T is a strong-mixing transformation on a probability space $(X, \mathcal{B}, \mu)$ then it is necessarily weak-mixing (and thus also ergodic).*

Proof. This is immediate from the definitions.

■

Example (Markov measures and shifts). Recall that a subshift of finite type $T : X \to X$ is defined on a space

$$X = \{x \in \prod_{n \in \mathbb{Z}} \{1, \dots, k\} : A(x_n, x_{n+1}) = 1, n \in \mathbb{Z}\}$$

for some $k \times k$ matrix A with entries either zero or unity. We define $T(x_n) = (x_{n+1})$ (i.e. all terms in the infinite sequences are shifted one place to the left). We shall assume in addition that A is aperiodic, i.e. there exists $n \geq 1$ such that for each $1 \leq i, j \leq k$ we have that $A^n(i, j) = 1$.

Let P denote a $k \times k$ stochastic matrix with entries $P(i, j) = 0$ iff $A(i, j) = 0$, and let p be its left eigenvector. Recall that we define the associated Markov measure by

$$\mu[i_0, \dots, i_{l-1}] = p(i_0)P(i_0, i_1) \dots P(i_{l-2}, i_{l-1}).$$

Let $C = [i_0, \ldots, i_{l-1}]$ and $D = [j_0, \ldots, j_{l-1}]$ be two cylinder sets. Observe that

$$C \cap T^{-(n+l)} D = \cup_{x_l, \ldots, x_{n+l-1}} [i_0, \ldots, i_{l-1}, x_l, \ldots, x_{n+l-1}, j_0, \ldots, j_{l-1}]$$

and thus

$$\begin{aligned}
&\mu(C \cap T^{-(n+l)} D) \\
&= \sum_{x_l, \ldots, x_{n+l-1}} p(i_0) P(i_0, i_1) \ldots P(i_{l-1}, x_l) \ldots P(x_{n+l-1}, j_0) \ldots P(j_{l-2}, j_{l-1}) \\
&= \mu(C)\mu(D) \left(\frac{1}{p(j_0)} \sum_{x_l, \ldots, x_{n+l-1}} P(i_{l-1}, x_l) \ldots P(x_{n+l-1}, j_0) \right) \\
&= \mu(C)\mu(D) \frac{P^n(i_{l-1}, j_0)}{p(j_0)}.
\end{aligned}$$

However, we know that $P^n(i_{l-1}, j_0) \to p(j_0)$ as $n \to +\infty$ (by writing P in terms of Jordan forms) and so we know that

$$\mu(C \cap T^{-(n+l)} D) \to \mu(C)\mu(D)$$

as $n \to +\infty$. For arbitrary sets $A, B \in \mathcal{B}$ we can cover them by unions of disjoint cylinders $A \subset \cup_{i=1}^n C_i$ and $B \subset \cup_{j=1}^m D_i$ such that $\mu((\cup_{i=1}^n C_i - C) < \epsilon$ and $\mu((\cup_{i=1}^m D_i - D) < \epsilon$ and then by approximation we see that $\mu(A \cap T^{-(n+l)} B) \to \mu(A)\mu(B)$ as $n \to +\infty$.

11.7 Comments and reference

We have given only the briefest introduction to the spectral theory associated with measure preserving transformations. A particularly nice introduction is contained in the appendix to [1].

Reference

1. W. Parry, *Topics in Ergodic Theory*, C.U.P., Cambridge, 1981.