

Mixing

1 Mixing conditions

Proposition 1.1 (Characterization of ergodicity). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space. A measure-preserving transformation $T: \mathcal{X} \rightarrow \mathcal{X}$ is ergodic if and only if, for every $A, B \in \mathcal{F}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k}B) = \mu(A)\mu(B) .$$

Proof. Apply Birkhoff's ergodic theory to the indicator of B , and then integrate over A . $\square$

Definition 1.2. Let $(\mathcal{X}, \mathcal{F}, \mu, T)$ be a probability space. A measure-preserving transformation $T: \mathcal{X} \rightarrow \mathcal{X}$ is said to be

- (strongly) mixing if for every $A, B \in \mathcal{F}$,

$$\lim_{n \rightarrow \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B) .$$

- weakly mixing if for every $A, B \in \mathcal{F}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} |\mu(A \cap T^{-k}B) - \mu(A)\mu(B)| = 0 .$$

Interpretation 1.3. Consider a point $x \in \mathcal{X}$ chosen at random according to μ . Strong mixing means that, for every measurable A and B , the events $x \in A$ and $T^n(x) \in B$ are asymptotically independent of each other. In other words, the system asymptotically forgets whether it started in A or outside A . $\diamond$

Clearly,

$$\text{strong mixing} \implies \text{weak mixing} \implies \text{ergodicity}.$$

The motivation for the concept of weak mixing will be clarified later.

Example 1.4 (Irrational rotation). Let α be an irrational, and $R_\alpha: \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ the rotation by α , i.e., $R_\alpha(x) := x + \alpha \pmod{1}$. As discussed in class, with respect to the Lebesgue measure on $\mathbb{R}/\mathbb{Z}$, R_α is ergodic but not strongly mixing. That R_α is not weakly mixing follows similarly from the characterization of weak mixing in terms of convergence in density (see below). $\circ$

Example 1.5 (Bernoulli shift). As shown in class, every Bernoulli shift is strongly mixing. $\circ$

The proof of the strong mixing property of Bernoulli shifts can be generalized to give the following:

Proposition 1.6 (Easier verification). *Let $T: \mathcal{X} \rightarrow \mathcal{X}$ be a measure-preserving map on a probability space $(\mathcal{X}, \mathcal{F}, \mu)$. Let $\mathcal{S}$ be a semi-algebra generating $\mathcal{F}$.*

- (a) *T is ergodic if and only if, for every $A, B \in \mathcal{S}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k}B) = \mu(A)\mu(B) .$$

(b) T is mixing if and only if, for every $A, B \in \mathcal{S}$,

$$\lim_{n \rightarrow \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B) .$$

(c) T is weakly mixing if and only if, for every $A, B \in \mathcal{S}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} |\mu(A \cap T^{-k}B) - \mu(A)\mu(B)| = 0 .$$

In other words, if $\mathcal{F}$ is generated by a family $\mathcal{S}$ of simple sets (e.g., intervals, or cylinders, ...) that form a semi-algebra, then in order to show that T is ergodic/mixing/weakly mixing, it is enough to verify the corresponding identity only for A and B that are simple.

Remark 1.7. Constructing an explicit example of a system that is weakly mixing but not strongly mixing is non-trivial. Nonetheless, classic results by Halmos and Rokhlin indicate that being weakly mixing but not strongly mixing is a generic property. More specifically, among all invertible measure-preserving transformations of a non-atomic standard probability space, those that are weakly but not strongly mixing form a residual set (with respect to the weak topology). See e.g., the paper by Steven Alpern (Inventiones Mathematicae, 1976). Concrete constructions of weakly but not strongly mixing systems can be found in the monograph by Petersen (1983). $\diamond$

Example 1.8 (Markov shift). Let Σ be a finite alphabet and $Q: \Sigma \times \Sigma \rightarrow [0, 1]$ a stochastic matrix. A probability distribution $p: \Sigma \rightarrow [0, 1]$ is stationary for the Markov chain with transition probabilities given by Q if $pQ = p$.

- (a) If Q is irreducible, then it has a unique stationary distribution p , and the Markov shift defined by the pair (Q, p) is ergodic.
- (b) If Q is irreducible and p its unique stationary distribution. The following are equivalent:
 - (i) Q is aperiodic.
 - (ii) The Markov shift defined by (Q, p) is weakly mixing.
 - (iii) The Markov shift defined by (Q, p) is strongly mixing.

The above statements essentially summarize the basics of finite-state Markov chain in the language of ergodic theory. $\circ$

2 More on weak mixing

2.1 Convergence in density

Natural density. The *upper* and the *lower density* of a set $J \subseteq \mathbb{N}$ are defined as

$$\begin{aligned} \overline{d}(J) &:= \limsup_{n \rightarrow \infty} \frac{|J \cap [0, n)|}{n} , \\ \underline{d}(J) &:= \liminf_{n \rightarrow \infty} \frac{|J \cap [0, n)|}{n} . \end{aligned}$$

If the upper and the lower density of J coincide, we call the common value the *density* of J and denote it by $d(J)$.

For instance,

$$d(\emptyset) = 0 , \quad d(\mathbb{N}) = 1 , \quad d(\{2k : k \in \mathbb{N}\}) = d(\{2k + 1 : k \in \mathbb{N}\}) = 1/2 .$$

The following is an example of a set which does not have a well-defined density.

Example 2.1 (A set without density). Each $J \subseteq \mathbb{N}$ can naturally be represented by its indicator sequence $x \in \{0, 1\}^{\mathbb{N}}$ where $x_i = 1$ if and only if $i \in J$. Conversely, every sequence $x \in \{0, 1\}^{\mathbb{N}}$ is the indicator sequence of one and only one set $J \subseteq \mathbb{N}$.

Consider the set J whose indicator sequence has the following form:

$$x_i := \begin{cases} 1 & \text{if } N_{2k} \leq i < N_{2k+1} \text{ for some } k \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

where $0 = N_0 < N_1 < N_2 < N_3 < \dots$ are such that $N_k/N_{k+1} \rightarrow 0$ as $k \rightarrow \infty$. (For instance, we can choose $N_k := 2^{k^2}$.) Then, it is easy to verify that $\underline{d}(J) = 0 \neq 1 = \bar{d}(J)$. $\circ$

We say that $J \subseteq \mathbb{N}$ has *full density* if $d(J) = 1$.

Remark 2.2. The natural density is somewhat like a probability measure on $\mathbb{N}$:

- (additivity) If $J_1 \cap J_2 = \emptyset$, then $d(J_1 \cup J_2) = d(J_1) + d(J_2)$ provided $d(J_1)$ and $d(J_2)$ exist.
- (monotonicity) If $J_1 \subseteq J_2$, then $d(J_1) \leq d(J_2)$ provided $d(J_1)$ and $d(J_2)$ exist.
- (sub-additivity) $d(J_1 \cup J_2) \leq d(J_1) + d(J_2)$ provided all three densities exist.

However, it is not countably additive, and lacks continuity:

- If $J_n := \{n\}$, then $d(\bigcup_{n=0}^{\infty} J_n) = 1 \neq 0 = \sum_{n=0}^{\infty} d(J_n)$.
- If $J_n := \{n, n+1, n+2, \dots\}$, then $d(\bigcap_{n=0}^{\infty} J_n) = 0 \neq 1 = \lim_{n \rightarrow \infty} d(J_n)$. $\diamond$

Full-density diagonal trick. Recall the *diagonal* trick often used in analysis: if

$$(a_n^{(0)})_{n \in \mathbb{N}}, (a_n^{(1)})_{n \in \mathbb{N}}, (a_n^{(2)})_{n \in \mathbb{N}}, \dots$$

is a sequence of sequences, each being a subsequence of the previous one, then there exists a sequence (the “diagonal” sequence) which is eventually a subsequence of each $(a_n^{(k)})_{n \in \mathbb{N}}$.

Clearly the diagonal trick is about the indices rather than the values of the sequences. The following is a concise formulation of the trick:

Lemma 2.3 (Diagonal sequence). *Let $\mathbb{N} = J_0 \supseteq J_1 \supseteq J_2 \supseteq \dots$ be a sequence of infinite sets. There exists an infinite set $J \subseteq \mathbb{N}$ (the “diagonal” set) such that $J \setminus J_k$ is finite for each $k \in \mathbb{N}$.*

Here, we should think of J_k as the set of indices of the k -th subsequence of $(a_n^{(0)})_{n \in \mathbb{N}}$.

It turns out that if all the sets J_k have full density, then the diagonal set can also be chosen to have full density.

Lemma 2.4 (Full-density diagonal sequence). *Let $\mathbb{N} = J_0 \supseteq J_1 \supseteq J_2 \supseteq \dots$ be a sequence of sets such that $d(J_k) = 1$ for each $k \in \mathbb{N}$. There exists a set $J \subseteq \mathbb{N}$ (the “diagonal” set) such that*

- (i) $J \setminus J_k$ is finite for each $k \in \mathbb{N}$, and
- (ii) $d(J) = 1$.

Proof. Pick $0 = N_0 < N_1 < N_2 < N_3 < \dots$ such that for each k , we have

$$\frac{|J_k \cap [0, n]|}{n} > 1 - 2^{-k} \quad \text{for all } n \geq N_k.$$

We construct J by pruning $\mathbb{N}$ as follows: for each k , we remove those indices $n \geq N_k$ that are not in J_k . More specifically, we let

$$J := \mathbb{N} \setminus \bigcup_{k=1}^{\infty} (J_k^c \cap [N_k, \infty)) = \mathbb{N} \cap \bigcap_{k=1}^{\infty} ([0, N_k) \cup J_k).$$

Then,

- (i) For each k , we have $J \setminus J_k \subseteq [0, N_k)$, hence $J \setminus J_k$ is finite.

- (ii) Note that if $n < N_{k+1}$, then $J \cap [0, n) \supseteq J_k \cap [0, n)$ because $J_k \subseteq J_{k-1} \subseteq \dots \subseteq J_0$. For each n , let k_n denote the unique value such that $N_{k_n} \leq n < N_{k_n+1}$. Then,

$$\frac{|J \cap [0, n)|}{n} \geq \frac{|J_{k_n} \cap [0, n)|}{n} > 1 - 2^{-k_n}.$$

Since $k_n \rightarrow \infty$ as $n \rightarrow \infty$, we find that

$$\underline{d}(J) = \liminf_{n \rightarrow \infty} \frac{|J \cap [0, n)|}{n} \geq \liminf_{n \rightarrow \infty} (1 - 2^{-k_n}) = 1.$$

Therefore, $d(J) = 1$. □

The following are two applications of the latter lemma. The proofs are left as exercises

Proposition 2.5 (Cesàro convergence vs. convergence in density). *Let $a_0, a_1, \dots \in \mathbb{R}$ be a sequence of real numbers belonging to a bounded interval $I = [C, D]$. The following are equivalent:*

- (i) (Cesàro convergence to an extremal value)

$$\frac{1}{n} \sum_{k=0}^{n-1} a_k \rightarrow C \text{ as } n \rightarrow \infty.$$

- (ii) (convergence in density)

There exists a $J \subseteq \mathbb{N}$ with $d(J) = 1$ such that $a_n \rightarrow C$ as $J \ni n \rightarrow \infty$.

Proposition 2.6 (Convergence in density vs. statistical convergence). *Let $a_0, a_1, \dots \in \mathbb{R}$ and $a \in \mathbb{R}$. The following are equivalent:*

- (i) (convergence in density)

There exists a $J \subseteq \mathbb{N}$ with $d(J) = 1$ such that $a_n \rightarrow a$ as $J \ni n \rightarrow \infty$.

- (ii) (statistical convergence)

$$\text{For every } \varepsilon > 0, \lim_{n \rightarrow \infty} \frac{1}{n} |\{k < n : |a_k - a| \geq \varepsilon\}| = 0.$$

As a corollary of the first proposition, we have the following elegant characterization of weak mixing.

Theorem 2.7 (Weak mixing in terms of convergence in density). *Let $T: \mathcal{X} \rightarrow \mathcal{X}$ be a measure-preserving map on a probability space $(\mathcal{X}, \mathcal{F}, \mu)$. The following conditions are equivalent:*

- (i) T is weakly mixing.

- (ii) For every $A, B \in \mathcal{F}$, there exists a set $J_{A,B} \subseteq \mathbb{N}$ with $d(J_{A,B}) = 1$ such that

$$\lim_{J_{A,B} \ni n \rightarrow \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Remark 2.8. Can the set $J_{A,B}$ be chosen independently of A and B ?

In the above characterizations, the condition $A, B \in \mathcal{F}$ can be relaxed to $A, B \in \mathcal{S}$ for a generating semi-algebra $\mathcal{S}$. Using full-density diagonalization, we find that the sets $J_{A,B}$ can be chosen uniformly in A and B as long as $\mathcal{F}$ has a countable generating set. ◇

Corollary 2.9. *Let $T: \mathcal{X} \rightarrow \mathcal{X}$ be a measure-preserving map on a probability space $(\mathcal{X}, \mathcal{F}, \mu)$. The following conditions are equivalent:*

- (i) T is weakly mixing.

- (ii) For every $A, B \in \mathcal{F}$,

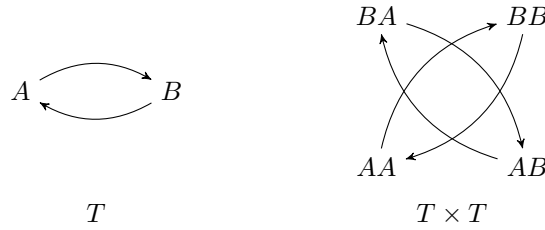
$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} |\mu(A \cap T^{-k}B) - \mu(A)\mu(B)|^2 = 0.$$

2.2 Product characterization

The (Cartesian) product of two dynamical systems T_1 and T_2 is a dynamical system that contains of independent copies of T_1 and T_2 that run simultaneously. More precisely, let $T_1: \mathcal{X}_1 \rightarrow \mathcal{X}_1$ and $T_2: \mathcal{X}_2 \rightarrow \mathcal{X}_2$ be two measure preserving dynamical systems. The product system is the system $T_1 \times T_2: \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathcal{X}_1 \times \mathcal{X}_2$ on $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$, where $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the product σ -algebra, $\mu_1 \times \mu_2$ is the product measure, and $(T_1 \times T_2)(x_1, x_2) := (T_1(x_1), T_2(x_2))$.

The product of two ergodic systems need not be ergodic.

Example 2.10. A finite dynamical system consisting of a single periodic orbit of period $\ell \geq 2$ is ergodic with respect to the uniform measure, but its product with itself has more than one periodic orbits and is hence not ergodic.



○

Exercise. Count the number of ergodic components of $T \times T$ if T is a finite system with a single periodic orbit of period ℓ .

Weak and strong mixing conditions are however closed under products.

Proposition 2.11.

- (a) The product of every two weakly mixing systems is again weakly mixing.
- (b) The product of every two strongly mixing systems is again strongly mixing.

Proof sketch. Using the convergence in density characterization of weak mixing, the proofs become very similar, so let us focus on strong mixing.

Let $T_1: \mathcal{X}_1 \rightarrow \mathcal{X}_1$ and $T_2: \mathcal{X}_2 \rightarrow \mathcal{X}_2$ be measure-preserving transformations of probability spaces $(\mathcal{X}_1, \mathcal{F}_1, \mu_1)$ and $(\mathcal{X}_2, \mathcal{F}_2, \mu_2)$. We need to show that, for every $A, B \in \mathcal{F}_1 \otimes \mathcal{F}_2$,

$$(\mu_1 \times \mu_2)(A \cap T^{-n}B) \rightarrow (\mu_1 \times \mu_2)(A)(\mu_1 \times \mu_2)(B) \quad \text{as } n \rightarrow \infty.$$

First, verify this for measurable rectangles $A = A_1 \times A_2$ and $B = B_1 \times B_2$. Recall that measurable rectangles form a semi-algebra generating the product σ -algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$. The claim thus follows from Proposition 1.6. $\square$

Interpretation 2.12. Example 2.10 illustrates that if two ergodic (but not weakly mixing) systems are initially “coupled”, they may remain “coupled” indefinitely. On the other hand, if two weakly/strongly mixing systems are initially “coupled”, they gradually become “decoupled” in the following sense.

More precisely, let $A \subseteq \mathcal{X}_1 \times \mathcal{X}_2$ be an arbitrary measurable set with $(\mu_1 \times \mu_2)(A) > 0$. Then, strong mixing implies that, for every measurable $B_1 \in \mathcal{X}_1$ and $B_2 \in \mathcal{X}_2$ and large n ,

$$(\mu_1 \times \mu_2)(B_1 \times B_2 | A) \approx \mu_1(B_1)\mu_2(B_2),$$

which means the possible initial codependence of the two systems is asymptotically forgotten. A similar statement holds for weak mixing for n outside a full-density set of time steps. $\diamond$

Weak mixing turns out to be the minimal condition that ensures the above property.

Theorem 2.13 (Weak mixing in terms of products). *Let $T: \mathcal{X} \rightarrow \mathcal{X}$ be a measure-preserving map on a probability space $(\mathcal{X}, \mathcal{F}, \mu)$. The following conditions are equivalent:*

- (i) T is weakly mixing.

(ii) $T \times T$ is weakly mixing.

(iii) $T \times T$ is ergodic.

Proof. The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) have already been established. Let us prove (iii) $\Rightarrow$ (i).

Suppose that $T \times T$ is ergodic. We use the characterization in Corollary 2.9 to show that T is weakly mixing. Let $A, B \in \mathcal{F}$. Observe that

- $\frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k})^2 = \frac{1}{n} \sum_{k=0}^{n-1} (\mu \times \mu)((A \times A) \cap (T \times T)^{-k}(B \times B)) \rightarrow \mu(A)^2 \mu(B)^2$ as $n \rightarrow \infty$.
- T is ergodic (why?), hence $\frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k} B) \rightarrow \mu(A) \mu(B)$ as $n \rightarrow \infty$.

It follows that

$$\begin{aligned}
 & \frac{1}{n} \sum_{k=0}^{n-1} \left[\mu(A \cap T^{-k} B) - \mu(A) \mu(B) \right]^2 \\
 &= \frac{1}{n} \sum_{k=0}^{n-1} \left[\mu(A \cap T^{-k})^2 - 2\mu(A) \mu(B) \mu(A \cap T^{-k} B) + \mu(A)^2 \mu(B)^2 \right] \\
 &= \frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k})^2 - 2\mu(A) \mu(B) \frac{1}{n} \sum_{k=0}^{n-1} \mu(A \cap T^{-k} B) + \mu(A)^2 \mu(B)^2 \\
 &\rightarrow \mu(A)^2 \mu(B)^2 - 2\mu(A) \mu(B) \mu(A) \mu(B) + \mu(A)^2 \mu(B)^2 \\
 &= 0.
 \end{aligned}$$

Hence T is weakly mixing. □

Recall that an irrational rotation on $\mathbb{R}/\mathbb{Z}$ is not weakly mixing. The following exercise generalizes this.

Exercise. Show that a non-trivial measure-preserving isometry of a metric space cannot be weakly mixing. More specifically, let $T: \mathcal{X} \rightarrow \mathcal{X}$ be an isometry of a metric space $\mathcal{X}$, and let μ be a probability measure on $\mathcal{X}$ that is T -invariant but not just a point mass at a fixed point of T . Show that T is not weakly mixing.

2.3 Spectral characterization