Rem. Recall: ergodicity was equiv. to

$$\text{(iv)} \quad \frac{1}{N}\sum_{n=1}^{N} T^n f \to \int f\,d\mu \cdot \mathbb{1} \text{ in } L^2 \quad \forall f \in L^2.$$

This is a version of the weak law of large numbers: For $f \in L^2$, consider $X_n := T^n f$ random variables on (X,μ), $n \in \mathbb{N}$, with $\int X_n = \int f\,d\mu$ $\forall n$.
Then for erg. systems,

$$\frac{X_1 + \dots + X_N}{N} \xrightarrow{L^2} \int f\,d\mu \cdot \mathbb{1}$$

(in stochastics: convergence in measure which is weaker)

3.3. Birkhoff's ergodic thm: the formulation

$\boxed{\text{Thm } 3.3}$ (Birkhoff)
Let (X,μ,T) be an MDS, $f \in L^1(X,\mu)$. Then

$$(*) \qquad \frac{1}{N}\sum_{n=1}^{N} (T^n f)(x)$$

conv. for a.e. x. If (X,μ,T) is ergodic, then the limit equals $\int f\,d\mu$ a.e.

Rem. 1) Thm 3.3 is called pointwise ergodic theorem

2) Birkhoff $\Rightarrow$ von Neumann and the limit of $(*)$ is the same for $f \in L^2$, i.e., $= P_{\mathrm{Fix}\,T} f$. In general the limit is $\mathbb{E}(f, \Sigma')$. In particular, if (X,μ,T) is ergodic, then the limit equals $\int f\,d\mu \cdot \mathbb{1}$.

exerc.

condit. expectation

σ-algebra of T-inv. sets

$\boxed{\text{Cor. } 3.4}$ An MDS (X,μ,T) is erg. $\Leftrightarrow$ $\forall f \in L^1$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} (T^n f)(x) = \int f\,d\mu \quad \text{a.e.},$$

i.e., „time mean = space mean" a.e.

(if $f = \mathbb{1}_A$, then $\lim \underbrace{\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_A(T^n x)}_{\#\{n \le N:\ T^n x \in A\}} = \mu(A)$)

proof $\Rightarrow$ Birkhoff

$\Leftarrow$ Let $f \in \mathrm{Fix}\,T$. By assumption,

$$f = \frac{1}{N}\sum_{n=1}^{N} T^n f \to \int f\,d\mu \cdot \mathbb{1} \quad \text{a.e.},$$

i.e., f const. a.e.

Rem. So Boltzmann's erg. hypothesis was proved only for ergodic systems and only a.e. Note that ergodicity is necessary for „space mean = time mean"

How to prove Birkhoff?

Observation On a dense subset

$$D := \operatorname{Fix} T \oplus (1-T)(L^\infty(X,\mu)) \subset L^1$$

the Cesàro means (ergodic means)

$$S_N f := \frac{1}{N} \sum_{n=1}^{N} T^n f$$

conv. a.e. (even in L^∞):

- $f \in \operatorname{Fix} T \Rightarrow S_N f = f \to f$
- $f = g - Tg$, $g \in L^\infty$, then

$$S_N f = \frac{1}{N} \sum_{n=1}^{N} (T^n g - T^{n+1} g) = \frac{Tg - T^{N+1} g}{N} \xrightarrow[N\to\infty]{} 0.$$

$\|\cdot\|_\infty \le 2\|g\|_\infty$

And we know: $\overline{D}^{\|\cdot\|_1} = L^1$ by von Neumann's decomp. and the density of L^∞ and L^2 in L^1

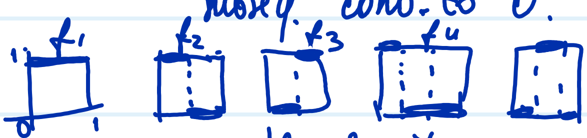
$\|\cdot\|_1 \le \|\cdot\|_2$ – CS

Problem: How to extend a.e. conv. from the dense set D to the whole L^1?

Rem.: A.e. convergence does not come from a topology: in a top./metric space:

$f_n \to f \iff \forall$ subsequence of (f_n) has a subseq. conv. to f

But $\exists (f_n) \subset L^\infty$ diverging in $\forall x$ but $\forall$ subseq. of (f_n) has a subseq. conv. to 0.



exer.

So: no simple density argument.

Idea: Find conditions which assure that the set

$$F := \{ f \in L^1 : (S_n f) \text{ conv. a.e.} \}$$

is closed in L^1. Since $\underset{\text{dense in } L^1}{D} \subset F$, this would imply $F = L^1$.

3.4. Maximal inequality and Banach's principle

Let (X,μ,T) be an MDS with Koopman operator T on L^1.

Def.

$$S_N f := \frac{1}{N} \sum_{n=0}^{N-1} T^n f$$

– convergence is the same as for $\frac{1}{N}\sum_{n=1}^{N}$ by the shift invariance of Cesàro limit

and

$$S^* f := \sup_{N\in\mathbb{N}} |S_N f| \quad \text{pointwise}$$

$$S^* : L^1 \to \left\{ \begin{array}{l} \text{measurable fcts with} \\ \text{values in } \mathbb{R} \cup \{+\infty\} \end{array} \right\}$$

– here σ-alg. is of the form A or $A \cup \{+\infty\}$, $A \in \Sigma$

– the maximal operator corresponding to (X,μ,T)

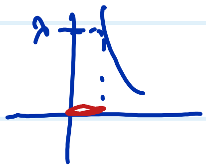
One can define max. operator for an arbitrary sequence (T_n) of operators on $L^1(X,\mu)$ by

$$\boxed{T^* f := \sup_{n\in\mathbb{N}} |T_n f|} \quad \text{(pointwise)}$$

Properties of T^*

- $T^* f \geq 0 \quad \forall f$
- $T^*(\alpha f) = |\alpha| \cdot T^* f \quad \forall f \in L^1, \forall \alpha \in \mathbb{C}$
- $T^*(f+g) \leq T^* f + T^* g \quad \forall f,g \in L^1$ – subadditivity

$\boxed{\text{Def}}$ A seq. of operators (T_n) on L^1 satisfies a (weak (1,1)) maximal inequality if



$$\mu(\{x : (T^* f)(x) > \lambda\}) \leq c(\lambda) \quad \begin{array}{l}\forall \lambda > 0 \\ \forall f \in L^1 \text{ with } \|f\|_1 \leq 1\end{array}$$

for a fct $c : (0,\infty) \to [0,\infty)$ with $\lim_{\lambda\to\infty} c(\lambda) = 0$

$\boxed{\text{Thm 3.5}}$ (Banach's principle)

Let (T_n) be a seq. of lin. bdd operators on $\underbrace{L^1(X,\mu)}_{\text{prob. space}}$. If the corresponding maximal operator T^* satisfies a maximal inequality, then the set

$$F := \{f \in L^1 : (T_n f) \text{ conv. a.e.}\}$$

is a closed subspace of L^1.

Proof subspace – clear. To show: closed.

Let $f \in \overline{F}$. We show: $(T_n f)$ is Cauchy a.e., i.e., $f \in F$

Def.
$$h := \overline{\lim_{k,l\to\infty}} |T_k f - T_l f| \quad \text{(pointwise)}$$

To show: $h = 0$ a.e.

For $g \in F$ observe:

$$|T_k f - T_l f| \leq \underbrace{|T_k(f-g)|}_{\leq T^*(f-g)} + |T_k g - T_l g| + \underbrace{|T_l(g-f)|}_{\leq T^*(f-g)}$$

$$\leq 2T^*(f-g) + |T_k g - T_l g|.$$

So
$$h \leq 2T^*(f-g) + \underbrace{\overline{\lim_{k,l\to\infty}} |T_k g - T_l g|}_{=0 \text{ by } g \in F} = 2T^*(f-g).$$

Let $\lambda > 0$. The maximal inequality implies (if $g \neq f$, but for $g = f$, $f \in F$ clear)

$$\mu[h > 2\lambda] \leq \mu[\underbrace{T^*(f-g) > \lambda}_{T^*\left(\frac{f-g}{\|f-g\|_1}\right) > \frac{\lambda}{\|f-g\|_1}}] \leq c\left(\frac{\lambda}{\|f-g\|_1}\right) \xrightarrow[g\to f]{} 0$$

So $\mu[h > 2\lambda] = 0 \quad \forall \lambda > 0$, so $h = 0$ a.e.
because $h \geq 0$

3.5 Maximal ergodic theorem and the rest of proof of Birkhoff's thm

For Birkhoff it suffices to show that the operators

$$S_N := \frac{1}{N} \sum_{n=0}^{N-1} T^n$$

satisfy a maximal ineq. (with $c(\lambda) := \frac{1}{\lambda}$)

$\boxed{\text{Thm 3.6}}$ (Maximal erg. thm)

Let (X, μ, T) be an MDS, $f \in L^1(X,\mu)$ real valued. Then $\forall N \in \mathbb{N}$

$$\int_{\left\{ \substack{x :\, f(x) + (Tf)(x) + \dots + (T^n f)(x) > 0 \\ \text{for some } n \leq N} \right\}} f \, d\mu \geq 0$$

in particular,

$$\int_{\left\{ \substack{x :\, f(x) + \dots + (T^n f)(x) > 0 \\ \text{for some } n} \right\}} f \, d\mu \geq 0$$

Proof: The last assertion follows from monotonicity of the sets

$$E_N := \left\{ x : \substack{f(x) + \dots + (T^n f)(x) > 0 \\ \text{for some } n \leq N} \right\} \nearrow$$

exer.

Let $N \in \mathbb{N}$ and def.

$$f_0 := 0$$
$$f_1 := f$$
$$\vdots$$
$$f_n := f + Tf + \dots + T^{n-1} f$$
$$F_N := \max\{f_0, \dots, f_N\}$$

We have $F_N \geq 0$ and $F_N \geq f_n \quad \forall n \in \{0, \dots, N\}$. So

$$T F_N + f \geq T f_n + f = f_{n+1} \qquad \forall n \in \{0, \dots, N\}$$

T is pos. (or just def. of T)

in particular,

$$T F_N + f \geq \max\{f_1, \dots, f_N\} \qquad (**)$$

Let $x \in E_N$, i.e., $F_N(x) > 0$. Then

$$0 < \max\{f_0, \dots, f_N\} = \max\{f_1, \dots, f_N\}$$

def. of F_N

and therefore

$$T F_N + f \geq F_N \text{ on } E_N$$

by $(**)$

Since F_N and also $T F_N \geq 0$, we have

$$\int_{E_N} f\,d\mu \geq \int_{E_N} F_N\,d\mu - \int_{E_N} TF_N\,d\mu = \underbrace{\int_X F_N\,d\mu}_{F_N=0 \text{ on } X\setminus E_N} - \int_{E_N} \underbrace{TF_N}_{\geq 0}\,d\mu$$

$$\geq \int_X F_N\,d\mu - \int_X TF_N\,d\mu \underset{T\ \mu\text{-pres.}}{=} 0 \qquad \blacksquare$$

As a corollary we have

$\boxed{\text{Thm 3.7}}$ (Maximal inequality)

Let (X,μ,T) be an MDS, $f\in L^1(X,\mu)$. Then

$$\mu[S^* f>\lambda] \leq \frac{\|f\|_1}{\lambda} \quad \forall \lambda>0.$$

In particular, S^* satisfies the maximal ineq. with $c(\lambda):=\frac{1}{\lambda}$, where

$$S^* f := \sup_n \left|\frac{1}{n}\sum_{j=0}^{n-1} T^n f\right|.$$

Proof: Consider $|f|-\lambda$ – real valued. The maximal erg. thm. implies

$$\int_{\{x:\ \exists n:\ \underbrace{|f|(x)+\dots+(T^{n-1}|f|)(x)-n\cdot\lambda>0}_{\frac{1}{n}\sum_{j=0}^{n-1}T^j|f|(x)>\lambda}\}} (|f|-\lambda)\,d\mu \geq 0,$$

i.e.,

$$\|f\|_1 \geq \int_{\{x:\ \frac{1}{n}\sum_{j=0}^{n-1}T^j|f|(x)>\lambda\}} |f|\,d\mu \underset{\text{by the above}}{\geq} \lambda\cdot\mu(\{x:\exists n: \underbrace{\frac{1}{n}\sum_{j=0}^{n-1}T^j|f|(x)}_{\geq\left|\frac{1}{n}\sum_{j=0}^{n-1}T^j f\right|(x)}>\lambda\})$$

$$\geq \lambda\cdot\mu(\{x \overset{\exists n}{:} \underbrace{\left|\frac{1}{n}\sum_{j=0}^{n-1}T^j f\right|(x)>\lambda}_{(T^*f)(x)>\lambda}\}) \qquad \blacksquare$$

Proof of Birkhoff's erg. thm (summary)

As in the proof of von Neumann,

$$S_N f := \frac{1}{N}\sum_{j=0}^{N-1} T^j f$$

conv. a.e. on

$$D := \underset{\text{conv. to } f\ (=f)}{\operatorname{Fix} T} \oplus \underbrace{(1-T)(L^\infty(X,\mu))}_{\text{conv. to } 0}$$

By von Neumann's decomp., D is dense in L^1.

By Thm. 3.7 (maximal ineq.), we have

$$\mu(\{x:(S^*f)(x)>\lambda\}) \leq \frac{\|f\|_1}{\lambda} \quad \forall\lambda>0,\ \forall f\in L^1$$

By Thm. 3.5 (Banach's principle) the set

$$F:=\{f\in L^1: (S_N f) \text{ conv. a.e.}\}$$

is closed in L^1. By $F\supset D$, F is also dense in $L^1 \Rightarrow = L^1$.

So $\frac{1}{N}\sum_{n=0}^{N-1} T^n f$ conv. a.e. $\forall f\in L^1$, and this is equiv. to conv. of $\frac{1}{N}\sum_{n=1}^{N} T^n f$ ■

3.6. Birkhoff's erg. theorem for uniquely erg. systems

Recall (Thm of Krylov–Bogolyubov)

$$\left.\begin{array}{l} X \text{ compact metric space} \\ T: X \to X \text{ cont} \end{array}\right\} \Rightarrow \exists \text{ T-inv. Borel prob. measure on } X$$

$\boxed{\text{Def. 3.8}}$ A topological dynamical system (X,T) is called **uniquely ergodic** if it has a unique T-inv. Borel prob. measure μ.

Rem. μ is automatically ergodic (if not, (completion) take $A \subset X$ T-inv. with $\mu(A) \in (0,1)$ and def. ν by $\nu(B) := \frac{\mu(B \cap A)}{\mu(A)}$ exerc. ν is T-inv., $\nu \neq \mu$)

More holds: $\underbrace{\mathrm{ext}}_{\text{extremal points}}(\underbrace{M_T(X)}_{\text{all T-inv. prob. meas.}}) = \{\text{erg. T-inv. prob. measures on } X\}$ see exerc. sheets

$\boxed{\text{Ex}}$ We saw before:

$$(\mathbb{T}, a) \text{ uniquely erg.} \iff a \text{ irrat.}$$

and in this case $\mu = $ Leb. rescaled

$\boxed{\text{Thm. 3.9}}$ (**Birkhoff's erg. theorem for uniquely erg. systems**)
Let (X,T) be uniquely erg. with T-inv. prob. measure μ. Then

$$\frac{1}{N}\sum_{n=1}^{N} f(T^n x) \longrightarrow \int f \, d\mu \qquad \forall f \in C(X) \ \forall x \in X$$

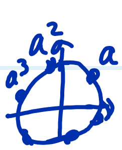
Proof let $x \in X$. As in the proof of Krylov-Bogolyubov:

$\forall$ limit point of $\left\{\frac{1}{N}\sum_{n=1}^{N} \delta_{T^n x}\right\}_{N=1}^{\infty}$ is a T-inv. prob. measure $\Rightarrow = \mu$ by unique erg.,

i.e.,
$$\frac{1}{N}\sum_{n=1}^{N} \delta_{T^n x} \longrightarrow \mu \quad \text{in } C(X)',$$

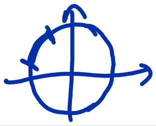
which means
$$\frac{1}{N}\sum_{n=1}^{N} f(T^n x) \to \int f \, d\mu \quad \forall f \in C(X).$$ ■

3.7. Application to equidistribution



Let $a \in \mathbb{T}$ be irrat. We know: $\{a^n : n \in \mathbb{N}\}$ is dense in $\mathbb{T}$. But much more is true

$\boxed{\text{Thm 3.10}}$ (**Weyl, equidistribution**)
The sequence $\{a^n : n \in \mathbb{N}\}$ is **equidistributed** in $\mathbb{T}$ for irrat. $a \in \mathbb{T}$, i.e.,



$\forall$ interval $I \subset \mathbb{T}$

$$\frac{\{n \in \{1,\dots,N\} : a^n \in I\}}{N} \longrightarrow \text{length}(I) \quad \text{rescaled}$$

Proof We know: $(\mathbb{T}, a)$ is uniquely erg. for irrat. a, so Birkhoff for uniquely erg. systems:

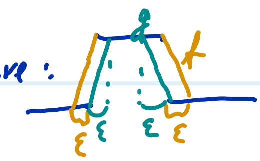
$$\frac{1}{N}\sum_{n=1}^{N} f(a^n) \to \int f \, d\mu \qquad \forall f \in C(\mathbb{T})$$

we took $x := 1$

We need it for $f := \mathbb{1}_I$ – not continuous

idea: approximation.

Let $\varepsilon \in (0,1)$ and def. continuous f, g as here:



with $g \le \mathbb{1}_I \le f$ and

$$\int f - \varepsilon \le \text{length}(I) \le \int g + \varepsilon$$

Then

$$\frac{1}{N}\sum_{n=1}^{N} g(a^n) \le \underbrace{\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_I(a^n)} \le \frac{1}{N}\sum_{n=1}^{N} f(a^n)$$

$\ell(I) - \varepsilon \le \int g$ (length) $\qquad \int f \le \ell(I) + \varepsilon$

$$\ell(I) - \varepsilon \le \underline{\lim}\, \underline{\quad} \le \overline{\lim}\, \underline{\quad} \le \ell(I) + \varepsilon$$

$\forall \varepsilon > 0$, so $\lim \underline{\quad} = \ell(I)$. ∎

Rem. Birkhoff's erg thm for general systems (recall: $(\mathbb{T}, a)$ is erg. for irrat. a) implies a somewhat different statement: $\forall A \subset \mathbb{T}$ measur.

$$\frac{\{n \in \{1,\dots,N\} : a^n x \in A\}}{N} \longrightarrow \mu(A)$$

for a.e. x

3.8 Further applications

① Normal numbers

$\boxed{\text{Def}}$ A number $x \in [0,1]$ is called simply normal (base 10) if in the decimal repr.

$$x = 0, x_1 x_2 x_3 \dots$$

the seq. (x_n) is equidistributed in $\{0,\dots,9\}$, i.e.,

$$\frac{|\{n \le N : x_n = k\}|}{N} \longrightarrow \frac{1}{10} \qquad \forall k \in \{0,\dots,9\}$$

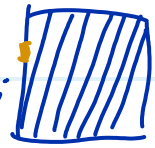
(each digit appears asympt. with prob. $\frac{1}{10}$)

$\boxed{\text{Ex}}$ $0{,}0123\dots90123\dots9\dots$

$\boxed{\text{Thm}}$ (Borel) A.e. $x \in [0,1]$ is simply normal (base 10)

Proof Recall: the decimal repr. is unique for a.e. x

Consider $Tx := 10x \mod 1$ on $[0,1)$



We know: $([0,1), \text{Leb}, T)$ is isomorphic to the Bernoulli shift $B(\frac{1}{10}, \dots, \frac{1}{10}) := (\{0,\dots 9\}^{\mathbb{N}}, \underset{\text{measure}}{\text{product}}, \leftarrow)$

via
$$\Phi : \{0,\dots,9\}^{\mathbb{N}} \to [0,1)$$
$$(x_1, x_2, \dots) \mapsto 0,x_1x_2\dots$$

Since $B(\frac{1}{10}, \dots, \frac{1}{10})$ is ergodic, so is $([0,1), \text{Leb}, T)$

exer.: $A \subset [0,1)$ is T-inv. $\iff \Phi^{-1}(A)$ is shift inv.

Let $k \in \{0,\dots 9\}$ be a digit. Consider
$$A := \{x \in [0,1) : x_1 = k\} = [\tfrac{k}{10}, \tfrac{k+1}{10})$$

We have $(A) = \frac{1}{10}$ and by Birkhoff (general systems)
$$\frac{\#\{n \in \{1,\dots,N\} : x_n = k\}}{N} = \frac{1}{N}\sum_{n=0}^{N-1} \mathbb{1}_A(T^n x) \to \mu(A) = \frac{1}{10}$$
for a.e. x ∎

Rem. A number is called <u>normal</u> if $\forall$ base ≥ 2 $\forall d$ $\forall$ combination of d digits appears asympt. with prob. $\frac{1}{b^d}$.

Exerc. As above one shows: a.e. number $\in [0,1)$ is normal.

$\boxed{\text{Ex}}$ $0{,}12345678910111213\dots$ – Champernowne's number – normal base 10, but all bases is open

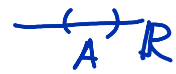
$\sqrt{2}, \pi, \ln 2, e$ – open (even whether $\forall$ digit comes ∞-often)

2 The strong law of large numbers

$\boxed{\text{Thm}}$ (Kolmogorov) Let $(\Omega, \mathbb{P})$ be a prob. space and $(X_n)_{n=1}^{\infty} \subset L^1(\Omega, \mathbb{P})$ be independent, identically distributed (i.i.d.) real random variables. Then
$$\boxed{\frac{X_1 + \dots + X_N}{N} \to \mathbb{E}(X_1) \quad \mathbb{P}\text{-a.e.}}$$

Rem. Recall: • identically distributed: $\mathbb{P}(X_n^{-1}(A))$ is indep. of n $\forall$ meas. A, i.e., ν given by
$$\nu(A) := \mathbb{P}(X_n^{-1}(A))$$
which is a prob. measure on $\mathbb{R}$ called the distribution of X_n is indep. of n.

• indep. if the sets $(X_n^{-1}(A_n))$ are indep., i.e.,

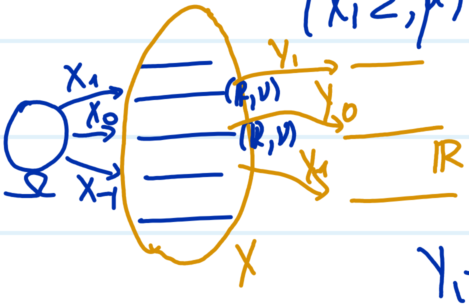


$\forall k$ $\forall A_{j_1}, \dots, A_{j_k}$
$$\mathbb{P}\Big(\bigcap_{m=1}^{k} X_{j_m}^{-1}(A_{j_m})\Big) \overset{\mathbb{P}(X_{j_1}^{-1}(A_{j_1}))\cdot \dots \cdot \mathbb{P}(X_{j_m}^{-1}(A_m))}{=} \nu(A_{j_1}) \cdot \ldots \cdot \nu(A_{j_k})$$

Proof using Birkhoff Consider $\mathbb{E}(X_1) = \int_\Omega \underbrace{X_1(\omega)}_{=:t} d\mathbb{P}(\omega) = \int_{\mathbb{R}} t \, d\nu(t)$ – expected value of $\forall X_n$

Def. a new prob. space

$$(X, \Sigma, \mu) := \Big(\mathbb{R}^{\mathbb{N}}, \underbrace{\prod_{n=1}^{\infty} \mathrm{Borel}(\mathbb{R})}_{\text{product } \sigma\text{-algebra}}, \prod_{n=1}^{\infty} \nu\Big)$$



Let $\tau : X \to X$ be the left shift and $Y_n : X \to \mathbb{R}$ be projection onto nth coord.

Then $Y_{j+1}(x) = Y_1(\tau^j x)$, so

$$\frac{Y_1 + \dots + Y_N}{N} = \frac{1}{N}\sum_{n=0}^{N-1} Y_1(\tau^n x) \xrightarrow[N\to\infty]{} \int_X Y_1 \, d\mu \quad \text{a.e.}$$

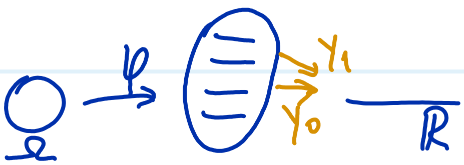
By Birkhoff because our system is mixing $\Rightarrow$ ergodic
the same proof as before for Bernoulli shifts

By $\underbrace{\mu(Y_1^{-1}(A)) = \nu(A)}_{\mu \text{ is a product measure}}$, $A \subset \mathbb{R}$, the limit equals

$$\int_X \underbrace{Y_1}_{=:t} \, d\mu = \int_{\mathbb{R}} t \, d\nu(t) = \mathbb{E}(X_1) \text{ — the right limit!}$$

Exer. $Y_1 \in L^1(X, \mu)$

But we need (X_j) and not (Y_j)



Consider $\varphi : \Omega \to X$ by

$$\varphi(\omega) := (X_n(\omega))_{n=1}^{\infty}$$

Claim φ is measure preserving, i.e., $\mathbb{P}(\varphi^{-1}(B)) = \mu(B)$ $\forall B \subset X$ meas.

$$\varphi^{-1}\Big(\text{cylinder generated by } A_{j_1}, \dots, A_{j_k}\Big) = \bigcap_{i=1}^{k} X_{j_i}^{-1}(A_{j_i}), \quad A_{j_i} \subset \mathbb{R}$$

$$\mathbb{P}(\varphi^{-1}(\text{cylinder} \dots)) = \mathbb{P}(\cap \dots) = \nu(A_{j_1}) \cdot \ldots \cdot \nu(A_{j_k}) = \mu(\text{cylinder} \dots)$$

(μ product measure)

Claim is proved because cylinders generate the product σ-alg. ■-claim

Moreover, we have $\boxed{X_n = Y_n \circ \varphi}$, so

$$\frac{X_1 + \dots + X_N}{N}(\omega) = \frac{Y_1 + \dots + Y_N}{N}(\varphi(\omega))$$

and

$$\Big\{\omega : \frac{X_1 + \dots + X_N}{N}(\omega) \not\to \mathbb{E}(X_1)\Big\} = \underbrace{\varphi^{-1}\underbrace{\Big\{x : \frac{Y_1 + \dots + Y_N}{N}(x) \not\to \mathbb{E}(X_1)\Big\}}_{\text{null set by Birkhoff}}}_{\text{null set since } \varphi \text{ is measure pres.}}$$

The proof is complete. ■

Rem. Analogously, the weak law of large numbers for i.i.d. variables follows from von Neumann's mean erg. thm. (exer.)

4. Strong and weak mixing

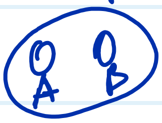
4.1. Strong mixing

Ex Mix wine and water. When are they mixed well enough?

wine | water

Intuitively: $\forall B$. $\dfrac{\mu(\text{wine in a region } B)}{\mu(B)} \approx$ proportion of wine in the whole glas

Def. 4.1 An MDS (X,μ,T) is called mixing (or strongly mixing) if
$$\mu(T^{-n}A\cap B)\xrightarrow[n\to\infty]{}\mu(A)\cdot\mu(B)\quad\forall A,B \text{ meas.}$$



Ex Bernoulli shifts are mixing, $([0,1], \times 2 \bmod 1)$ also since isom. to Bernoulli $(\{0,1\}^{\mathbb{N}}, \Pi, \leftarrow)$

Rem. (X,μ,T) is mixing $\Rightarrow$ ergodic

Recall: if A is T-inv., $\mu(\underbrace{T^{-n}A}_{A \text{ up to a null set}}\cap A)\to\mu(A)^2$, so $\mu(A)=\mu(A)^2 \Rightarrow \mu(A)\in\{0,1\}$

Ex (erg. $\not\Rightarrow$ mixing)

$X=\{0,1\}$, $\mu=(\frac12,\frac12)$, T ↻

Then (X,μ,T) is erg. ($\{0\},\{1\}$ not inv.) but not mixing.
$$\mu(T^{-n}\{0\}\cap\{0\})=\begin{cases}1/2, & 2\mid n\\ 0 & \text{otherwise},\end{cases}$$
so it diverges

Thm 4.2 (Characterisation of mixing)
For an MDS (X,μ,T) with Koopman op. T on L^2 TFAE:

(i) (X,μ,T) is strongly mixing

(ii) $\mu(T^{-n}A\cap A)\to\mu(A)^2$ $\forall A$ meas.

(iii) $T^n f\to\int f\,d\mu\cdot\mathbb{1}$ weakly $\forall f\in L^2$, i.e.,
$$\langle T^n f,g\rangle\to\int f\,d\mu\cdot\int\bar g\,d\mu\quad\forall f,g\in L^2$$

in particular, in this case
$$L^2(X,\mu)=\mathbb{C}\cdot\mathbb{1}\overset{\perp}{\oplus}\{f:\ T^n f\to 0 \text{ weakly}\}$$

Proof The last assertion follows from the trivial decomp
$$L^2(X,\mu)=\mathbb{C}\cdot\mathbb{1}\overset{\perp}{\oplus}\{f:\ \underbrace{f\perp\mathbb{1}}_{\int f\,d\mu=0}\}$$
and (iii).

(i)$\Rightarrow$(ii) clear

(iii)$\Rightarrow$(i) Take $f:=\mathbb{1}_A$, $g:=\mathbb{1}_B$

(ii)$\Rightarrow$(iii) Let $A\in\Sigma$ and take $f:=T^k\mathbb{1}_A$, $g:=T^l\mathbb{1}_A$ for $k,l\in\mathbb{N}_0$
Then for every $n>l-k$
$$\langle T^n f,g\rangle=\langle T^{n+k}\mathbb{1}_A, T^l\mathbb{1}_A\rangle\overset{T \text{ is isom.}}{=}\langle T^{n+k-l}\mathbb{1}_A,\mathbb{1}_A\rangle$$
$$=\mu(T^{-(n+k-l)}A\cap A)\xrightarrow[n\to\infty]{\text{by (ii)}}\mu(A)^2=\int f\cdot\int\bar g$$

Since $\langle T^n f,g\rangle\to\int f\int\bar g$ also if $f=\mathbb{1}$ or $g=\mathbb{1}$, it also holds $\forall f,g\in \operatorname{lin}\{\mathbb{1}, T^k\mathbb{1}_A, k\geq 0\}=:Y$ – T-inv. lin. subspace of $L^2(X,\mu)$ for the Koopman op. T

By approximation (use $\|T^n\|\leq 1$ $\forall n$), we have conv. $\forall f,g\in\overline{Y}$ (exerc.)

Let now $f\in\overline{Y}$, $g\in\overline{Y}^\perp$. Then
$$\langle \underbrace{T^n f}_{\in\overline{Y}},\underbrace{g}_{\perp\overline{Y}}\rangle=0=\int f\cdot\underbrace{\int\bar g\,d\mu}_{g\perp\overline{Y} \text{ and } \mathbb{1}\in Y},$$
so we have conv. $\forall g\in L^2$ and $\forall f\in\overline{Y}$.

Since A was arbitrary (meas.), we have $T^n f \to \int f \, d\mu \cdot \mathbb{1}$ $\forall$ simple fct $\Rightarrow$ $\forall f \in L^2$ by approx. (why?) ■

Ex (Rotations are not mixing)

Cons. $(\mathbb{T}, a)$.

Case 1: a rat. $\Rightarrow$ $(\mathbb{T}, a)$ not erg. $\Rightarrow$ not mixing

Case 2: a irrat. (so $(\mathbb{T}, \text{Leb}, a)$ is ergodic)

Consider $f(z) := z$, $f: \mathbb{T} \to \mathbb{T}$. Then

$$(Tf)(z) = a \cdot z = a \cdot f(z) \quad \forall z,$$

i.e., $Tf = a \cdot f$ implying $\langle \underbrace{T^n f}_{a^n f}, f \rangle = a^n \cdot \|f\|^2 \underset{n\to\infty}{\not\to} 0 = \underbrace{\langle f, \mathbb{1} \rangle}_{=0} \cdot \langle \mathbb{1}, \bar{f} \rangle$,

so $(\mathbb{T}, a)$ not mixing. More is true:

$$L^2(\mathbb{T}, \mu) = \overline{\text{lin}} \{e_n\}_{n=-\infty}^{\infty}$$

for $e_n(z) := z^n$ and $\forall e_n$ is an eigenfct of T, so

$$\langle T^k e_n, e_n \rangle = a^{kn} \text{ diverges in } k \text{ if } n \neq 0.$$

$(T^k e_n)(z) = e_n(a^k z) = a^{kn} z^n = a^{kn} e_n(z)$

Such a decomp. of L^2 into eigenspaces (with eigenfcts being called "characters", multiplicative fcts $G \to \mathbb{T}$) holds $\forall$ rotation on a comp. group (no proof).

So rotations have "opposite" behaviour to mixing.

We almost proved:

Prop. 4.3 (X, μ, T) is mixing $\Rightarrow$ $\begin{cases} P\sigma(T) \cap \mathbb{T} = \{1\} & \text{, point spectrum of } T \text{ (eigenvalues)} \\ 1 \text{ is a simple eigenvalue,} \\ \quad \text{i.e., } \ker(1-T) = \mathbb{C} \cdot \mathbb{1}. \end{cases}$

Proof As above, if λ is an eigenvalue with eigenfct f, then

$$\langle T^n f, f \rangle = \lambda^n \underbrace{\|f\|^2}_{\neq 0} \text{ diverges if } \lambda \neq 1.$$

if $\lambda = 1$: Assume $f \perp \mathbb{1}$, $Tf = f$. If T is mixing, then

$$\langle T^n f, f \rangle = \|f\|^2 \longrightarrow \underbrace{\langle f, \mathbb{1} \rangle}_{=0} \cdot \langle \mathbb{1}, \bar{f} \rangle = 0,$$

i.e., $f = 0$. So $\ker(1-T) = \mathbb{C} \cdot \mathbb{1}$ (since otherwise take $f \in \ker(1-T)$, $f \not\equiv$ const, then $f - \int f \perp \mathbb{1}$, also in $\ker(1-T)$) ■

Rem. $\not\Leftarrow$ as we will see soon

Observation Let $k \in \mathbb{N}$. Then

$$(X, \mu, T) \text{ is mixing} \iff (X, \mu, \underset{\text{also meas. pres.}}{T^k}) \text{ is mixing}$$

(False for ergodicity: $\circlearrowleft$)

Proof Let $f \in L^2$. By Thm. 4.2 we need to show:

$$T^n f \underset{n\to\infty}{\longrightarrow} 0 \text{ weakly} \iff T^{kn} f \underset{n\to\infty}{\longrightarrow} 0 \text{ weakly} \quad ((\Leftarrow) \; f \perp \mathbb{1})$$

$(\Rightarrow)$ clear

$(\Leftarrow)$ Let $j \in \{0, \dots, k-1\}$, $g \in L^2$. Then

$$\langle T^{kn+j} f, g \rangle = \langle T^{kn} f, \underbrace{(T^j)^* g}_{\text{fixed}} \rangle \underset{n\to\infty}{\longrightarrow} 0,$$

so $\langle T^m f, g \rangle \underset{m\to\infty}{\longrightarrow} 0$ (why?) ■

$k \quad 2k \quad 3k$

4.2. Weak mixing

since T is isometry

Question Does $\begin{cases} P_\sigma(T) \cap \cancel{\mathbb{T}} = \{1\} \\ 1 \text{ simple eigenvalue} \end{cases}$ imply strong mixing?

if not, which asymptotical property does it imply?

$\boxed{\text{Def}}$ For a set $\mathcal{J} \subset \mathbb{N}$, the **density** of $\mathcal{J}$ is def by

$\{x \;.\; x\,x\;.\;x\}$. $\quad d(\mathcal{J}) := \lim_{N\to\infty} \frac{|\mathcal{J} \cap \{1,\dots,N\}|}{N}$

N

if lim exists. if not, def. **upper** and **lower density** $\bar{d}$ and $\underline{d}$ by $\overline{\lim}$ and $\underline{\lim}$, respectively.

Rem. $d(\mathcal{J}) \in [0,1]$, $d(\mathcal{J}) = 1$ means that $\mathcal{J}$ is asympt. almost $\mathbb{N}$.

$\boxed{\text{Def. 4.4}}$ An MDS (X,μ,T) is called **weakly mixing** if $\forall A, B \in \Sigma$ $\exists \mathcal{J} \subset \mathbb{N}$ with $d(\mathcal{J}) = 1$ s.t.

$$\mu(T^{-n}A \cap B) \xrightarrow[\substack{n\to\infty \\ n\in\mathcal{J}}]{} \mu(A)\mu(B)$$

We write: $D\text{-}\lim \mu(T^{-n}A\cap B) = \mu(A)\mu(B)$

D for density 1

Unnatural, but easy to check (mixing: natural, very diff. to check)

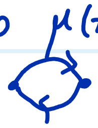
Rem. Mixing $\underset{\not\Leftarrow (2)}{\overset{(1)}{\Rightarrow}}$ weak mixing $\underset{\not\Leftarrow (4)}{\overset{(3)}{\Rightarrow}}$ ergodicity

(1) clear (take $\mathcal{J} := \mathbb{N}$)

(3) Assume $A \subset X$ is meas., T-inv. Then w. mixing implies

$$\mu(A) = \mu(T^{-n}A \cap A) \xrightarrow[\substack{n\to\infty \\ n\in\mathcal{J}}]{} \mu(A)^2,$$

so $\mu(A) = \mu(A)^2$, i.e., $\mu(A) \in \{0,1\}$.

(4)  erg., $\quad \mu(T^{-n}\{0\} \cap \{0\}) = \begin{cases} \frac{1}{2}, & 2 \mid n \\ 0 & \text{otherw.} \end{cases} \xrightarrow[\text{along any subseq.}]{} \hspace{-2em}\not\hspace{2em} \mu(A)\mu(A) = \frac{1}{4}$

(2) Very difficult

typical in Baire category sense

1944: Halmos "A generic transformation is ~~mixing~~ weak mixing"

1948: Rokhlin " —"— not mixing" — strong mixing

$([0,1], \text{Leb})$, $\mathcal{T} := \{T : [0,1] \to [0,1] \text{ meas. pres.}\}$

complete metric space for strong conv. of Koopman op.:

$T_n \to T$ if $T_n f \to Tf$ in L^2 $\forall f$.

(or: $\mu(T_n^{-1}A \,\Delta\, T^{-1}A) \to 0 \;\forall A$)

generic property = complement of 1. category.

So: a „typical" transf. is weakly but not strongly mixing.

The first concrete example: Chacon 1969

$\boxed{\text{Lemma 4.5}}$ (**Koopman – von Neumann**)

Let $(a_n) \subset [0,\infty)$ be a seq. of positive bounded numbers. Then

$$\frac{1}{N}\sum_{n=1}^{N} a_n \to 0 \iff D\text{-}\lim a_n = 0.$$

<u>Proof</u> ($\Leftarrow$) Let $C := \sup a_n$, $\varepsilon > 0$. Then by ass. $\exists n_\varepsilon$: $\forall n > n_\varepsilon,\ n \in J,\ 0 \le a_n < \varepsilon$. For such $N > n_\varepsilon$

$$\frac{1}{N}\sum_{n=1}^{N} a_n \le \underbrace{\frac{1}{N}\sum_{n=n_\varepsilon+1}^{N} a_n}_{\underbrace{\frac{1}{N}\sum_{\substack{n=n_\varepsilon+1\\ n\notin J}}^{N} a_n}_{\le \frac{C}{N}\cdot|\{1,\dots,N\}\setminus J|} + \frac{1}{N}\sum_{\substack{n=n_\varepsilon+1\\ n\in J}}^{N} a_n < \varepsilon} + \frac{n_\varepsilon \cdot C}{N} \le C\Big(\underbrace{1 - \frac{|\{1,\dots,N\}\cap J|}{N}}_{\to 0 \text{ by } d(J)=1}\Big) + \underbrace{\varepsilon \cdot \frac{(N-n_\varepsilon)}{N}}_{<\varepsilon} + \underbrace{\frac{n_\varepsilon C}{N}}_{\xrightarrow{N\to\infty} 0}$$

$< 3\varepsilon$ for N large enough.

($\Rightarrow$) (Sketch)

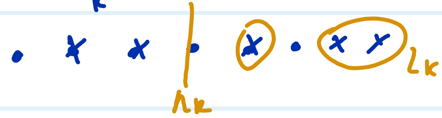
For $k \in \mathbb{N}$ def. $L_k := \{n : a_n \ge \frac{1}{k}\}$

Then $L_1 \subset L_2 \subset L_3 \subset \dots$ and $d(L_k) = 0\ \forall k$ by

$$\frac{|L_k \cap \{1,\dots,N\}|}{N} \underset{\substack{\uparrow\\ 1 \le k\cdot a_n \text{ for } n\in L_k}}{\le} \frac{1}{N}\sum_{n=1}^{N} k a_n = k \cdot \frac{1}{N}\sum_{n=1}^{N} a_n \xrightarrow{N\to\infty} 0 \text{ by assumpt.}$$

Let n_k be such that $\frac{|L_k \cap \{1,\dots,N\}|}{N} < \frac{1}{k}$ $\forall N > n_k$, WLOG $n_k \nearrow$

Def. $L := \bigcup_{k=1}^{\infty} L_k \cap [n_k, \infty)$



- $d(L) = 0$:

  For $m > n_1$ $\exists k$: $n_k \le m < n_{k+1}$, so

  $$L \cap \{1,\dots,m\} \underset{\text{since } L_1,\dots,L_{k-1} \subset L_k}{=} L_k \cap \{1,\dots,m\}$$

  and

  $$\frac{|L \cap \{1,\dots,m\}|}{m} = \frac{|L_k \cap \{1,\dots,m\}|}{m} \underset{\substack{\uparrow\\ m \ge n_k}}{\le} \frac{1}{k} \xrightarrow{m\to\infty} 0$$

- $a_n \xrightarrow[n\notin L]{n\to\infty} 0$: indeed, for $n \notin L$, $n \in [n_k, n_{k+1})$ one has $n \notin L_k$, i.e., $a_n < \frac{1}{k}$ ∎

$\boxed{\text{Thm. 4.6}}$ (<u>Characterisation of weak mixing</u>)

Let (X, μ, T) be an MDS with Koopman op. T on L^2. TFAE:

(i) T is $\boxed{\text{weakly mixing}}$

(ii) $\forall A, B \in \Sigma$
$$\frac{1}{N}\sum_{n=1}^{N} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| \xrightarrow{N\to\infty} 0$$

(iii) $\forall f, g \in L^2$
$$\frac{1}{N}\sum_{n=1}^{N} |\langle T^n f, g\rangle - \int f \cdot \int \bar{g}| \xrightarrow{N\to\infty} 0$$

(iv) $\boxed{P_\sigma(T) = \{1\} \text{ and } 1 \text{ is a simple eigenvalue}}$.

In this case $\boxed{L^2(X,\mu) = \mathbb{C}\cdot 1 \oplus^{\perp} \{f : T^n f \xrightarrow[n\in J]{n\to\infty} 0 \text{ weakly for some } J \subset \mathbb{N} \text{ with } d(J) = 1\}}$

<u>Rem.</u> Functions from ⟶ are called <u>weakly mixing fcts.</u>

Proof (i) $\Leftrightarrow$ (ii) Koopman–von Neumann for $a_n := |\mu(T^{-n}A\cap B) - \mu(A)\mu(B)|$

(ii) $\Rightarrow$ (iii) For charact. fcts clear $\rightsquigarrow$ simple fcts $\rightsquigarrow$ all fcts by approx. (exerc.)

(iii) $\Rightarrow$ (ii) clear: $f := \mathbb{1}_A$, $g := \mathbb{1}_B$

(iii) $\Rightarrow$ (iv) Let f satisf. $Tf = \lambda f$. Then $\lambda \in \mathbb{T}$ (T isom.)

Case 1: $\lambda \neq 1$. Then

$$\langle f, \mathbb{1}\rangle \underset{T\text{ isom.}}{=} \langle \underbrace{Tf}_{\lambda f}, \underbrace{T\mathbb{1}}_{\mathbb{1}}\rangle = \underset{\neq 1}{\lambda} \cdot \langle f, \mathbb{1}\rangle, \text{ so } f \perp \mathbb{1}.$$

By (iii) for $g := f$

$$\frac{1}{N}\sum_{n=1}^{N} \underbrace{|\langle T^n f, f\rangle|}_{|\lambda^n|\cdot\|f\|^2} = \|f\|^2 \longrightarrow 0, \text{ i.e., } f = 0.$$

Case 2 $\lambda = 1$.

(iii): $$\frac{1}{N}\sum_{n=1}^{N} \big|\underbrace{\langle T^n f, f\rangle}_{\|f\|^2} - |\textstyle\int f|^2\big| \longrightarrow 0,$$

i.e., $\|f\|^2 = |\int f|^2$, so $\|f\| = |\int f| = |\langle f, \mathbb{1}\rangle|$

Equality in the Cauchy-Schwarz ineq.: $f = c\cdot\mathbb{1}$ for some $c \in \mathbb{C}$, so 1 is simple eigenvalue.

(iv) $\Rightarrow$ (iii) (most difficult direction)

Sketch We show this direction under the

Assumption: (X, μ, T) is invertible