CHAPTER I BASIC DEFINITIONS AND CONSTRUCTIONS

I.1 MOTIVATION

1.1 LAW OF THE LARGE NUMBERS

$X_1, X_2, \dots : \Omega \to \mathbb{R}$ iid random variables $(\mathbb{E}[|X_1|] < \infty)$,

then

- $$\lim_{n\to\infty} \mathbb{P}\left[\left| \frac{1}{n} \sum_{k=1}^{n} X_k - \mathbb{E}[X_1] \right| > \varepsilon \right] = 0 \quad \forall \varepsilon > 0$$ (weak)
- $$\mathbb{P}\left[\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \mathbb{E}[X_1] \right] = 1$$ (strong)

Is independence necessary? → No!

1.2 A SIMPLE WEATHER MODEL

X_i : i^{th} day's weather $\in \{R, S\}$ "rainy"/"sunny"

$\mathbb{P}[X_{i+1} = b \mid X_i = a] = P_{ab} \in \mathbb{R}^{2\times 2}$, $a, b \in \{R, S\}$

Let

$$P = \begin{array}{cc} \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} & \begin{array}{c} R \\ S \end{array} \\ \begin{array}{cc} R & \quad S \end{array} & \end{array}$$

$\{X_i\}_{i \geq 0}$ is a Markov Chain

$$\mathbb{P}_S[X_i = S] = [0 \;\; 1] P^i \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

↑ "$X_0 = S$"

Eigendecomposition: $M = \begin{bmatrix} \overset{v^T}{\boxed{1/3 \quad 2/3}} \\ \underset{w^T}{\boxed{-1/2 \quad 1/2}} \end{bmatrix}$, $MPM^{-1} = \begin{bmatrix} \overset{\lambda_1}{\enclose{circle}{1}} & \\ & \underset{\lambda_2}{\enclose{circle}{0.4}} \end{bmatrix}$

v, w : left eigenvectors

1.3 LONG-TERM BEHAVIOR

$$\mathbb{E}\left[\overbrace{\frac{\#\text{sunny days in the first } n \text{ days}}{n}}^{=: W_n^S \text{ for sunny},\ W_n^R \text{ for rainy}}\right] =$$

$$= \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{P}_f[X_i = S] \qquad f \in \mathbb{R}^2 \text{: initial distribution}$$

$$= \frac{1}{n}\sum_{i=0}^{n-1} f^T P^i \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= f^T M^{-1} \frac{1}{n} \sum_{i=0}^{n-1} \begin{bmatrix} 1 & \\ & 0.4 \end{bmatrix}^i M \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\xrightarrow[n\to\infty]{} f^T \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2/3 = \nu_2 = \nu_{\text{"sunny"}}$$

Observations:

(a) $\mathbb{E}[W_n^{S/R}] \xrightarrow[n\to\infty]{} \nu_{\text{sunny/rainy}}$

regardless the initial condition

(b) It holds even ("strong law"):

$$W_n^{S/R} \xrightarrow[n\to\infty]{a.s.} \nu_{\text{sunny/rainy}}$$

(c) $f \in \mathbb{R}^2_{\geq 0}, \|f\|_1 = 1$: $\quad f^T P^n \xrightarrow[n\to\infty]{} \nu^T \qquad$ ("limiting distribution")

Is there something similar for deterministic systems as well?

1.4 DYNAMICAL SYSTEMS

Let $(G, +)$ be an (additive) group or semigroup, X some set, and $(T^g)_{g\in G}$ a family of maps $T^g : X \to X$, such that

(i) $T^{g_1} \circ T^{g_2} = T^{g_1+g_2}$ $\quad \forall g_{1,2} \in G$

(ii) $T^0 = id$

Then $(X, T^\bullet)$ is called a dynamical system with state space X, and $T^\bullet$ is called the flow.

- $G = \mathbb{N}_0$ or $\mathbb{Z}$: discrete time system, here $T := T^1$
- $G = \mathbb{R}_{\geq 0}$ or $\mathbb{R}$: continuous time system

1.5 THE STUDY OF DYNAMICAL SYSTEMS

(a) Differentiable dynamics : T differentiable, X smooth manifold

Local stretching ($\| DT \|$), Lyapunov exponents, ...

$\rightsquigarrow$ see eg. [Ma]

(b) Topological dynamics : T continuous, X topological space

E.g. "transitivity": U, V open sets; is there a $n \geq 0$ such that

$$T^{-n}(U) \cap V \neq \emptyset \; ?$$

$\rightsquigarrow$ see e.g. [BS]

(c) Ergodic theory : T measurable, X measure space

E.g. long-term behavior as above

1.6 PIONEERS

~ 1890 — Poincaré : "What is the fate of the solar system?"

~ 1870 — Boltzmann : "Macroscopic properties of a gas made from many interacting atoms evolving deterministically?"

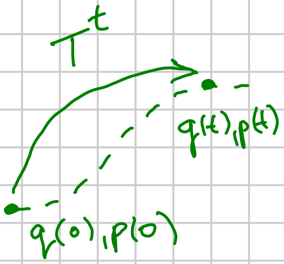
→ planets, atoms, ...

Classical mechanics: N objects with position $q_i \in \mathbb{R}^3$ and momenta $p_i \in \mathbb{R}^3$. The total energy (potential + kinetic) is given by the Hamiltonian $H(q_1, \dots, q_N, p_1, \dots, p_N)$.

Law of motion:

$$\dot{q}_i(t) = \frac{\partial H}{\partial p_i}(q(t), p(t))$$

$$\dot{p}_i(t) = -\frac{\partial H}{\partial q_i}(q(t), p(t))$$



Flow:

$$T^t : (q(0), p(0)) \longmapsto (q(t), p(t))$$

$$X = \{ H(q,p) = \text{const} \} \subseteq \mathbb{R}^{6N} \quad (\text{usually bounded})$$

How does the system behave for large t? Especially: For which (q,p) will $T^t(q,p)$ be close to (q,p) for some $t > 0$ (large)?

1.7 PREVIEW

Poincaré's Recurrence Theorem (PRT)

For almost all $x = (q,p) \in X$ (measured by the Lebesgue measure), the system starting at x will return arbitrary close to x at arbitrary large times.

Question: What other dynamical information can be extracted from measure-theoretic properties of a system?

The Ergodic Hypothesis

For certain measures μ, many functions $f: X \to \mathbb{R}$, and many states $x \in X$:

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t f(T^s x)\, ds = \frac{1}{\mu(X)} \int_X f(x)\, d\mu(x) = \text{"}E_\mu[f]\text{"}$$

"time-average = space-average"; cf. law of the large numbers

Note: the ergodic hypothesis is a quantitative version of the PRT: with $f = \chi_{B_\varepsilon(x)}$ (characteristic function of an ε-ball) we get the relative time the system spends in $B_\varepsilon(x)$.

I.2 THE SETUP OF ERGODIC THEORY

1.8 MEASURE PRESERVING TRANSFORMATIONS

Def A: A measure space is a triplet $(X, \mathcal{B}, \mu)$ where

(i) X is a set (also called "space")

(ii) $\mathcal{B}$ is a σ-algebra (colection of subsets of X containing $\emptyset$ and being closed under complements and countable unions)

$A \in \mathcal{B}$ is called measurable

(iii) $\mu : \mathcal{B} \to [0, \infty]$ is a measure, i.e. a σ-additive function on $\mathcal{B}$.

$\mu(X) = 1 \leadsto$ "probability space"

Def B: A measure preserving transformation (mpt) is a quartet $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a measure space, and

(i) T is measurable, i.e. $E \in \mathcal{B} \Longrightarrow T^{-1}E := T^{-1}(E) \in \mathcal{B}$

(ii) μ is invariant, i.e. $\mu(T^{-1}E) = \mu(E) \quad \forall E \in \mathcal{B}$

A probability preserving transformation (ppt) is a mpt on a prob. space.

Ex's:

1) $X = S^1$ (circle with unit circumference), $Tx = x + \alpha \bmod 1$, where $\alpha \in \mathbb{R}$, μ is the Lebesgue measure

2) $X=\mathbb{R}^{6N}$, T^t the flow of a Hamilton system, $\beta \geq 0$:

$\mu(dx) := \exp(-\beta H(x))\,dx$, i.e.

$$\mu(E) = \int_E \exp(-\beta H(x))\,dx$$

"Gibbs measure"

If μ exists it is invariant. In particular, for $\beta = 0$ we have that the Lebesgue measure is invariant (Liouville's theorem)

Def C: Two mpt's $(X_i, \mathcal{B}_i, \mu_i, T_i)$ are called measure-theoretically isomorphic if there is a measurable map $\pi : X_1 \to X_2$:

(i) there are $X_i' \in \mathcal{B}_i$ s.t. $\mu_i(X_i \setminus X_i') = 0$ and $\pi : X_1' \to X_2'$ is invertible with measurable inverse ("essentially a bijection")

(ii) for every $E \in \mathcal{B}_2$, $\pi^{-1}(E) \in \mathcal{B}_1$ and $\mu_1(\pi^{-1}E) = \mu_2(E)$

(iii) $T_2 \circ \pi = \pi \circ T_1$ on X_1

$$\begin{tikzcd} X_1 \arrow[r, "T_1"] \arrow[d, "\pi"'] & X_1 \arrow[d, "\pi"] \\ X_2 \arrow[r, "T_2"] & X_2 \end{tikzcd}$$

diagram commutes

1.9 POINCARÉ'S RECURRENCE THEOREM

Thm: Let $(X, \mathcal{B}, \mu, T)$ be a ppt. If $E \in \mathcal{B}$ with $\mu(E) > 0$ then for almost every $x \in E$ there is a sequence $n_k \to \infty$ s.t. $T^{n_k}(x) \in E$.

Proof ([BG, Thm 3.2.1], extended)

Let $N_0 \subseteq E$ the set of points that never return to E, i.e.

$$N_0 := \{ x \in E \mid T^k(x) \notin E, \ k \geq 1 \}.$$

(a) We show that $T^{-i}(N_0) \cap T^{-j}(N_0) = \emptyset \quad \forall\ i > j \geq 0$. If $x \in X$ s.t. $T^i(x) \in N_0 \ni T^j(x)$, then

$$T^{i-j}(\underset{\substack{\in \\ N_0}}{T^j(x)}) = T^i(x) \in N_0 \implies T^j(x) \in N_0 \text{ is recurrent } ↯$$

(b) N_0 is measurable, since $E \in \mathcal{A}$, T measurable, and

$$N_0 = \bigcap_{k=1}^{\infty} T^{-k}(E^c)$$

From ppt assumption & (b) $\Rightarrow \mu(T^{-i}(N_0)) = \mu(N_0) \ \forall i \geq 0$

Thus, from (a) $\Rightarrow$

$$1 = \mu(X) \geq \mu\left(\bigcup_{k=0}^{\infty} T^{-k}(N_0)\right) \overset{(a)}{=} \sum_{k=0}^{\infty} \mu(T^{-k}(N_0))$$

$$= \sum_{k=0}^{\infty} \mu(N_0)$$

Hence, $\mu(N_0) = 0$, and thus almost every point in E returns at least once to E: let $E_1 := E \setminus N_0$. Let N_i be the set that never returns to E_i, $i \geq 1$, and $E_{i+1} := E_i \setminus N_i$. We show iteratively that $\mu(N_i) = 0 \ \forall i \geq 0$.

The recurrent set, we want, is

$$E_\infty := E \setminus \bigcup_{i=0}^{\infty} N_i,$$

clearly measurable, and

$$\mu(E) \geq \mu(E_\infty) = \mu(E) - \mu\left(\bigcup_{i=0}^{\infty} N_i\right) \geq \mu(E) - \underbrace{\sum_{i=0}^{\infty} \mu(N_i)}_{=0}$$

$\blacksquare$

1.10 REMARKS TO THE PRT

(a) The finiteness of μ is essential, as the proof shows. E.g. take $X = \mathbb{Z}$, $\mathcal{B} = 2^X$, $T(x) = x+1$, and

$$\mu(E) = |E|, \text{ the counting measure}$$

Then, μ is invariant, but recurrence is clearly violated.

(b) $X = [0,1]$, $T(x) = 10x \pmod 1$

T leaves the Lebesgue measure invariant

PRT $\Longrightarrow$ for almost all $X \ni x = 0.\varepsilon_1\varepsilon_2\varepsilon_3 \ldots \varepsilon_n \ldots$ the sequence $\varepsilon_1\varepsilon_2 \ldots \varepsilon_n$ appears infinitely many times in the decimal expansion.

1.11 THE PROBABILISTIC POINT OF VIEW

- X is a sample space, $\omega \in X$ are possible states of a random system
- $\mathcal{B}$ is the collection of measurable events, i.e. $E \in \mathcal{B}$ are exactly the sets where we can decide whether $\omega \in E$.
- μ is the probability law: $\mathbb{P}[\omega \in E] = \mu(E)$
- measurable functions $f: X \to \mathbb{R}$ are random variables $f(\omega)$
- $X_n := f \circ T^n$ $(n \geq 1)$ is a stochastic process

  Its distribution:

$$\mathbb{P}\left[X_{i_1} \in E_{i_1}, \ldots, X_{i_k} \in E_{i_k}\right] = \mu\left(\bigcap_{j=1}^{k} \left\{\omega \in X \,\middle|\, f(T^{i_j}\omega) \in E_{i_j}\right\}\right)$$

Invariance of μ $\longrightarrow$ stationarity of (X_n):

$$\mathbb{P}\left[X_{i_1+m} \in E_{i_1}, \ldots, X_{i_k+m} \in E_{i_k}\right] = \mathbb{P}\left[X_{i_1} \in E_{i_1}, \ldots, X_{i_k} \in E_{i_k}\right] \quad \forall m \geq 0$$

Gain: bring in ideas/tools/intuition from probability theory; and vica versa: study stochastic phenomena in a completely deterministic manner.

1.12 INVARIANT SETS

Def: $E \in \mathcal{B}$ is called invariant, if $T^{-1}(E) = E$.

Note: If E is invariant:

(i) so is $E^c := X \setminus E$;

(ii) $T: E \to E$ and $T: E^c \to E^c$ can be considered separately, they do not interact.

Recall the ergodic hypothesis: $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(T^k x) = \frac{1}{\mu(X)}\int_X f\,d\mu$

Take $f = \chi_E$ and $x \in E^c$. The hypothesis tells us, if E^c is significant ($\mu(E^c) > 0$)

$$0 = \frac{\mu(E)}{\mu(X)} \Rightarrow \mu(E) = 0$$

↪ If E^c is significant, then E isn't: we can't decompose X into measure-theoretically significant invariant sets

1.13 ERGODICITY

Def: A mpt $(X, \mathcal{B}, \mu, T)$ is ergodic, if every invariant set E satisfies $\mu(E) = 0$ or $\mu(E^c) = 0$. We call μ an ergodic measure.

"The system is essentially indecomposable"

Thm: For a mpt the following are equivalent:

(a) μ is ergodic

(b) if $f: X \to \mathbb{R}$ is measurable and $f \circ T = f$ a.e., then $\exists c \in \mathbb{R}$ s.t. $f = c$ a.e.

(c) if $E \in \mathcal{B}$ and $\mu(T^{-1}E \,\Delta\, E) = 0$, then $\mu(E) = 0$ or $\mu(E^c) = 0$

Remarks:

- $A \Delta B := (A \setminus B) \cup (B \setminus A)$ is called the symmetric difference
- A set E with $\mu(T^{-1}E \,\Delta\, E) = 0$ is called essentially T-invariant

- Part (b) of the Thm can be read as "a.e. trajectory of T goes almost everywhere in X"

Proof: We prove (a)$\Leftrightarrow$(b), for the rest see [Sa, Prop. 1.1].

"(b) $\Rightarrow$ (a)": Let $E\in\mathcal{B}$ be an invariant set. Then $f=\chi_E$ satisfies

$$f\circ T = \chi_E\circ T = \chi_{T^{-1}E} = \chi_E = f,$$

hence $f=$ const a.e. Thus $\mu(E)=0$ or $\mu(E^c)=0$, and ergodicity follows.

"$\neg$(b) $\Rightarrow \neg$(a)": Let f violate (b), i.e. $f=f\circ T$ but f not constant a.e. Then $\exists\, c\in\mathbb{R}$ s.t. if $E:=\{f\geq c\}$ (and $E^c=\{f<c\}$) then $\mu(E)>0$ and $\mu(E^c)>0$. But

$$T^{-1}E=\{x\in X \mid f(Tx)\geq c\} \underset{\substack{\uparrow \\ f\circ T=f \text{ a.e.}}}{=} \{x\in X \mid f(x)\geq c\}=E,$$

so E is essentially invariant, and the mpt fails to be ergodic. ■

1.14 ERGODICITY: INVARIANT UNDER MEASURE-THEORETIC ISOMORPHISM

Prop: Assume that the mpts $(X_i,\mathcal{B}_i,\mu_i,T_i)$, $i=1,2$, are meas. th. isomorph. Then, if one is ergodic, so is the other.

Proof: Suppose the mpt #1 is ergodic.
Let $f: X_2\to\mathbb{R}$ be measurable and s.t. $f\circ T_2=f$ a.e.
Then

$$\begin{array}{l} f\circ T_2\circ\pi = f\circ\pi \text{ a.e.} \\ \quad\shortparallel \\ f\circ\pi\circ T_1, \text{ so by Thm 1.13: } f\circ\pi=\text{const a.e.}\end{array}$$

Since π is an essential bijection, $f=$ const a.e.

$\xrightarrow{\text{Thm 1.13}}$ mpt #2 is ergodic ■

1.15 INDEPENDENCE AND MIXING

- $(X, \mathcal{B}, \mu)$ prob. space
- $E \in \mathcal{B}$: events
- $\mathbb{P}[x \in E] = \mu(E)$: probability law

$E, F \in \mathcal{B}$ are independent, if

$$\mu(E) = \underbrace{\mu(E|F)}_{\text{conditional prob.}} := \frac{\mu(E \cap F)}{\mu(F)} \iff \mu(E \cap F) = \mu(E)\mu(F)$$

Def: A ppt $(X, \mathcal{B}, \mu, T)$ is called (strongly) mixing if for every $E, F \in \mathcal{B}$ we have $\mu(E \cap T^{-k}F) \xrightarrow[k \to \infty]{} \mu(E)\mu(F)$

"$T^{-k}(F)$ is asymptotically indepent of E" or

"$T^{k}(x)$ looses dependence on x for large k"

Exercise: Mixing is an invariant under measure-theoretic isom.

PropA: Strong mixing implies ergodicity.

Proof: $E \in \mathcal{B}$ invariant: $E = T^{-k}(E)$ $\forall k \geq 0$

$$\rightsquigarrow \mu(E) = \mu(E \cap T^{-k}E) \xrightarrow[k \to \infty]{} \mu(E)^2 \Rightarrow \mu(E) \in \{0, \overset{\overset{\mu(X)}{\|}}{1}\}$$

■

PropB: Strong mixing $\iff$ $\forall f, g \in L^2(X, \mu)$:

$$\int f \cdot g \circ T^k \, d\mu \xrightarrow[k \to \infty]{} \int f \, d\mu \cdot \int g \, d\mu$$

Proof: "$\Leftarrow$": Take $f = \chi_E$, $g = \chi_F$

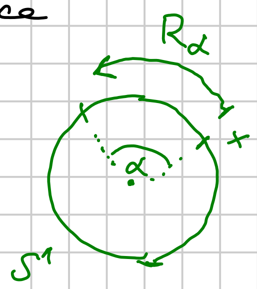
"$\Rightarrow$": [Sa, Prop 1.3]

■

I.3 EXAMPLES

1.16 CIRCLE ROTATION

$\alpha \in \mathbb{R}$: $R_\alpha : S^1 \to S^1$, $x \mapsto x+\alpha \mod 1$ $\quad S^1$: circle with unit circumference



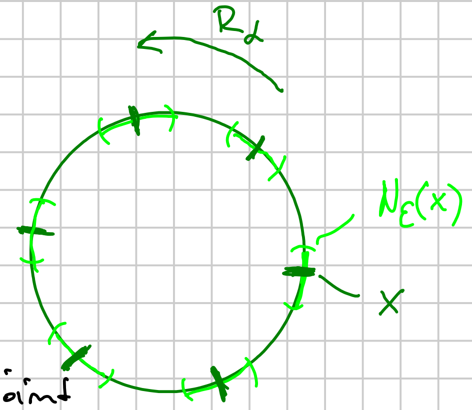
$\mathcal{B}$: Borel σ-alg, $\mu = m$: Lebesgue measure

Prop: (a) R_α is measure preserving

(b) R_α is ergodic iff $\alpha \notin \mathbb{Q}$

(c) R_α is never mixing

Proof: (a) ⇝ Exercise



(b) • $\alpha = \frac{p}{q} \in \mathbb{Q}$ $\quad (p,q \in \mathbb{N})$

$\Rightarrow R_\alpha^q = id$

Pick $0<\varepsilon< 1/2q$

$\Rightarrow N_\varepsilon(x), N_\varepsilon(x+\frac{1}{q}), \dots, N_\varepsilon(x+\frac{q-1}{q})$ disjoint

$\uparrow$ "ε-neighborhood of x"

$\Rightarrow E = \bigcup_{i=1}^{q-1} N_\varepsilon(x + k\frac{1}{q})$ is invariant and $0<m(E)<1$

$\Rightarrow R_\alpha$ not ergodic

• $\alpha \notin \mathbb{Q}$. Let $f = \chi_E$, E invariant $\Rightarrow f \circ R_\alpha = f$

Fourier series: $f(x) = \sum_{n\in\mathbb{Z}} c_n e^{2\pi i n x}$ (in L^2)

$$\rightsquigarrow \sum_{n\in\mathbb{Z}} c_n e^{2\pi i n x} = f(x) = f(R_\alpha x) = \sum_{n\in\mathbb{Z}} c_n e^{2\pi i n(x+\alpha)}$$

$$\Leftrightarrow c_n = 0 \vee e^{2\pi i n \alpha} = 1 \quad \forall n \neq 0$$

$$\Leftrightarrow c_n = 0 \;\forall n \neq 0 \Rightarrow f = \text{const} \Rightarrow m(E) \in \{0,1\}$$

$\Rightarrow R_\alpha$ ergodic

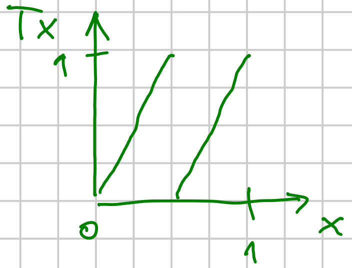
(c) Set $E=(0,\varepsilon)=F$, $0<\varepsilon<1$

Dirichlet thm: $\forall \alpha\in\mathbb{R}\ \exists n_k \xrightarrow[k\to\infty]{} \infty$ s.t. $n_k\alpha \mod 1 \to 0$ $(k\to\infty)$

$$\Rightarrow m(E\cap R_\alpha^{-n_k}F) \xrightarrow[k\to\infty]{} m(E)=\varepsilon \neq \varepsilon^2 = m(E)\,m(F)$$

$\Rightarrow R_\alpha$ not mixing

1.17 ANGLE DOUBLING



$X=S^1$, $\mu=m$, $T(x)=2x \mod 1$

Prop: T is measure preserving and strongly mixing

Proof: Meas. pres: Binary expansion (a.e. unique!)

$$x=0.d_1d_2\dots, \quad d_i\in\{0,1\}$$

Cylinder:

$$[\underline{d}]=[d_1,\dots,d_n]:=\{x\in S^1 \mid x=0.d_1d_2\dots d_n{*}{*}\dots\}$$

↑ whatever

$[\underline{d}]$ is interval of length 2^{-n}

$$T^{-1}[\underline{d}]=[\underbrace{*d_1d_2\dots d_n}_{n+1}], \quad *\in\{0,1\}$$

$$\Rightarrow m(T^{-1}[\underline{d}]) = m([0\underline{d}])+m([1\underline{d}])=2\cdot 2^{-(n+1)}=m([\underline{d}])$$

$\Rightarrow T$ meas. pres. on cylinders

Consider algebra $\mathcal{A}$ generated by $\{E\in\mathcal{B} \mid E=\bigcup_{i=1}^{k}[\underline{d}^i],\ [\underline{d}^i] \text{ cylinder}\}$,

and use the Monotone class thm to prove T meas. pres.

↗ Exercise

Mixing: $A=[\underline{a}]=[a_1,\dots,a_n]$, $B=[\underline{b}]=[b_1,\dots,b_m]$

$$\Rightarrow T^{-k}B=[\underbrace{*,\dots,*}_{k},b_1,\dots,b_m]$$

$$\Rightarrow A \cap T^{-\ell} B = [\underbrace{a_1,\dots,a_m}_{\underline{a}}, \underbrace{*,\dots,*}_{\ell-m}, \underline{b}] \quad \text{for } \ell \geq m$$

$$\Rightarrow m(A \cap T^{-k} B) = m([\underline{a}])\, m([\underline{b}]) \qquad \circledast$$

Cylinders generate the Borel σ-alg. $\mathcal{B}$

$\Rightarrow \forall E \in \mathcal{B} \ \exists \{A_\ell\} \subset \mathcal{A} : (A_\ell \triangle E) \downarrow \emptyset$ and $m(A_\ell) \xrightarrow[\ell\to\infty]{} m(E)$

$\circledast$ holds for the algebra gen. by the cylinders

$$\Rightarrow \quad m(A_\ell \cap T^{-\ell}[\underline{b}]) \xrightarrow[\ell\to\infty]{} m(E \cap T^{-\ell}[\underline{b}])$$

$$\|$$

$$m(A_\ell)\, m([\underline{b}]) \xrightarrow[\ell\to\infty]{} m(E)\, m([\underline{b}])$$

Do the same steps for $F \in \mathcal{B}$ to replace $[\underline{b}]$.

$\Rightarrow$ T mixing, hence ergodic.

1.18 BERNOULLI SCHEMES

- S — alphabet (finite set)
- $X = S^{\mathbb{N}} \ni \{x_k\} = (x_0, x_1, x_2, \dots)$
- $d(\{x_k\}, \{y_k\}) = 2^{-\min\{k \mid x_k \neq y_k\}}$ — metric

  ⤷ topology generated by cylinders

  $$[a_0, a_1, \dots, a_{n-1}] = \{\{x_k\} \mid x_i = a_i \quad 0 \leq i \leq n-1\}$$

  Cylinders generate the Borel σ-alg w.r.t. metric d.

- $T : X \to X,\ (x_0, x_1, x_2, \dots) \longmapsto (x_1, x_2, \dots)$ — left shift
- $\underline{p} = (p_a)_{a \in S} \in \mathbb{R}^{|S|}$ — probability vector (i.e. $\underline{p} \geq 0$, $\sum_a p_a = 1$)

Def: The Bernoulli measure corresponding to p on $\mathcal{B}$ (Borel σ-alg. wrt d) is the unique measure μ, s.t.

$$\mu[a_0,\dots,a_{n-1}] = \prod_{i=0}^{n-1} p_{a_i}$$

Prop A: The $(\frac{1}{2},\frac{1}{2})$-Bernoulli shift (i.e. $X=\{0,1\}$, $p=(\frac{1}{2},\frac{1}{2})$, μ the corresp. Bernoulli measure, T the left shift) is measure theoretically isomorphic to the angle doubling map.

Proof: Let $\pi(x_0,x_1,\dots) = \sum_{i=0}^{\infty} 2^{-(i+1)} x_i$

$\pi : X' \to S^1$ is bijective, where

$$X' = \{\{x_k\}\in X \mid \nexists\, n\in\mathbb{N} \text{ s.t. } x_\ell = 1 \ \forall \ell \geq n\}$$

and $\mu(X')=\mu(X)=1$ (exercise)

Clearly $\pi\circ T = \tilde{T}\circ\pi$ with $\tilde{T}: S^1\to S^1, x\mapsto 2x \bmod 1$

Check that $m(\pi[a_0,\dots,a_{n-1}]) = 2^{-n} = \mu[a_0,\dots,a_{n-1}]$

$$\xRightarrow{\text{Mon. class Thm.}} \pi \text{ preserves the measure} \xRightarrow{\text{Def. 1.8}} \text{claim}$$

↑ cylinders span $\mathcal{B}(X)$
+
dyadic intervals span $\mathcal{B}(S^1)$

∎

Prop B: Every Bernoulli scheme is mixing, hence ergodic.

Proof: The proof goes as in the case of the angle doubling map. For the $(\frac{1}{2},\frac{1}{2})$-Bernoulli shift we could use Prop 1.17 + Prop A (here) + the invariance of mixing under isomorphism to get the claim.

∎

$\left(\frac{1}{2},\frac{1}{2}\right)$-Bernoulli shift Coin toss

$\{x_k\} \in X$ sequence of events $Y_0, Y_1, \dots$ $Y_i \in \{0, 1\}$

heads tails

$\mu[a_0,\dots,a_{n-1}] = 2^{-n}$ $\mathbb{P}[Y_0 = a_0, \dots, Y_{n-1} = a_{n-1}] = 2^{-n}$

Invariance: $\mu = \mu \circ T^{-1}$

$\Rightarrow \mu[a_0] = \mu[*, a_0]$

$$\mathbb{P}[Y_0 = a_0] = \mathbb{P}[Y_0 = *, Y_1 = a_0] = \mathbb{P}[Y_1 = a_0]$$

independence

1.19 SUBSHIFT OF FINITE TYPE

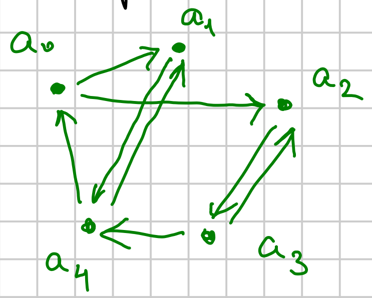
S – alphabet (finite), $A = \{t_{ij}\}_{i,j \in S} \in \{0,1\}^{|S|\times|S|}$

Def: The subshift of finite type (sft) with alphabet S and transition matrix A is:

- $\Sigma_A^+ = \left\{ \{x_k\} \in S^{\mathbb{N}} \mid t_{x_i x_{i+1}} = 1 \ \forall i \geq 0 \right\}$, and
- metric $d(\{x_k\}, \{y_k\}) = 2^{-\min\{k \mid x_k \neq y_k\}}$, and
- the action $T(x_0, x_1, x_2, \dots) = (x_1, x_2, \dots)$, the left shift

The sft is a compact metric space, and $T: \Sigma_A^+ \to \Sigma_A^+$ is continuous

Idea:



$a_i \in S$, $i = 1, \dots, |S|$

Edge from a_i to a_j iff $t_{a_i a_j} = 1$

$\Rightarrow \Sigma_A^+$: the set of all possible paths on this graph

Further terminology:

- Stochastic matrix: $P = (p_{ab})_{a,b \in S} \in [0,1]^{|S|\times|S|}$, $P\mathbb{1} = \mathbb{1} = (1,\dots,1)^T$

compatible with A: $t_{ab} = 0 \Rightarrow p_{ab} = 0$

- Probability vector: $p = (p_a)_{a \in S} \in [0,1]^{|S|}$ s.t. $\sum p_a = 1$
- Stationary probability vector: p s.t. $p^T P = p^T$

1.20 MARKOV SHIFT

Def: Given: prob. vector p, stoch. matrix P compatible with A.
The Markov measure on Σ_A^+ (with the Borel σ-alg $\mathcal{B}(\Sigma_A^+)$) is defined through

$$\mu[a_0, a_1, \dots, a_{n-1}] = p_{a_0} \cdot P_{a_0 a_1} \cdot \ldots \cdot P_{a_{n-2} a_{n-1}}$$

(Remark: P stochastic $\Rightarrow$ μ is well-defined Borel prob. measure)

Then, $(\Sigma_A^+, \mathcal{B}(\Sigma_A^+), \mu, T)$ is called the (p,P)-Markov shift.
↑ left shift

Prop: (a) p stationary wrt P $\Longleftrightarrow$ (p,P)-Markov shift is a ppt

(b) Every stoch. matrix has a stat. prob. vector

Proof: (a) Sufficient to show for cylinders: $\mu[\underline{b}] = \mu(T^{-1}[\underline{b}]) = \mu[*, \underline{b}]$:

$$\sum_a p_a P_{ab_0} \cdot \underline{P_{b_0 b_1}} \cdot \ldots \cdot \underline{P_{b_{n-2} b_{n-1}}} = p_{b_0} \underline{P_{b_0 b_1}} \cdot \ldots \cdot \underline{P_{b_{n-2} b_{n-1}}}$$

no division by 0 due to compatibility

$$\Longleftrightarrow \sum_a p_a P_{ab_0} = p_{b_0}$$

(b) Let $\Delta = \{ x \in \mathbb{R}^{|S|} \mid x \geq 0, \sum x_i = 1 \}$, and

$$R : \Delta \to \mathbb{R}^{|S|}, \quad x \longmapsto (x^T P)^T = P^T x$$

Check that $R(\Delta) \subseteq \Delta$ due to stochasticity of P

Brouwer's fixed point theorem (continuous mapping on a convex, closed set has a fixed point) $\Longrightarrow$ R has a fixpt ∎

Ergodicity & mixing of Markov shifts ⟶ later

I.4 BASIC CONSTRUCTIONS

1.21 PRODUCTS

$(X_i, \mathcal{B}_i, \mu_i)$, $i=1,2$, measure spaces

$(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_1 \times \mu_2)$ — product (measure) space

- $\mathcal{B}_1 \otimes \mathcal{B}_2 = \sigma\left(\{B_1 \times B_2 \mid B_i \in \mathcal{B}_i\}\right)$
- $(\mu_1 \times \mu_2)(B_1 \times B_2) := \mu_1(B_1)\mu_2(B_2)$ (unique!)

Def: The product of two mpts $(X_i, \mathcal{B}_i, \mu_i, T_i)$, $i=1,2$, is the mpt $(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_1 \times \mu_2, T_1 \times T_2)$ with

$$(T_1 \times T_2)(x_1, x_2) = (T_1 x_1, T_2 x_2)$$

Prop: The product of two ergodic mpts is not necessarily ergodic. The product of two mixing mpts is always mixing.

Proof: The product of two (identical) ergodic circle rotations:

$$\underset{R_\alpha \times R_\alpha}{\overset{\parallel}{T}} : S^1 \times S^1 \to S^1 \times S^1, \quad (x_1, x_2) \longmapsto (x_1 + \alpha, x_2 + \alpha) \bmod 1$$

is not ergodic: $f(x,y) := x - y \bmod 1$ is a non-constant invariant function (cf. Thm. 1.13)

Mixing: see [Sa, Prop. 1.9]

1.22 SKEW-PRODUCTS

Example: $(\Sigma, \mathcal{B}, \mu, \sigma)$: $(\frac{1}{2}, \frac{1}{2})$-Bernoulli shift $(\Sigma = \{0,1\}^{\mathbb{N}})$

$f: \Sigma \to \mathbb{Z}$, $f(\{x_k\}) = (-1)^{x_0}$

Let: $T: \Sigma \times \mathbb{Z} \to \Sigma \times \mathbb{Z}$, $T(\{x_k\}, \ell) = (\sigma\{x_k\}, \ell + f(\{x_k\}))$

T is mpt wrt $\mu \times m_{\mathbb{Z}}$ ($m_{\mathbb{Z}}$: counting measure), and

$$T^n(\{x_k\}, \ell) = (\sigma^n\{x_k\}, \ell + Y_0 + Y_1 + \dots + Y_{n-1}), \quad Y_i := (-1)^{x_i}$$

"random walk on $\mathbb{Z}$ driven by the noise process $(X, \mathcal{B}, \mu, \sigma)$"

Def: Let $(\Omega, \mathcal{F}, \rho, \sigma)$ be a mpt, $(X, \mathcal{B}, \mu)$ a prob. space, and $\{T_\omega\}_{\omega \in \Omega}$ a family of mpts on $(X, \mathcal{B}, \mu)$. Define the skew-product as $(\Omega \times X, \mathcal{F} \otimes \mathcal{B}, \rho \times \mu, \tau)$ with

$$\tau(\omega, x) = (\sigma\omega, T_\omega x)$$

τ preserves $\rho \times \mu$ ⇝ exercise.

$$T^n(\omega, x) = (\sigma^n \omega, T_{\sigma^{n-1}\omega} \circ T_{\sigma^{n-2}\omega} \circ \dots \circ T_{\sigma\omega} \circ T_\omega x)$$

⇝ skew-products model forced / driven / time-dependent / random dyn. systems

The skew-product is sometimes called a random dynamical system

$(\Omega, \mathcal{F}, \rho, \sigma)$ is also called base or driving system

Example: T_ω: change in the amount of ice in a glacier

ω: season ($\sigma^n \omega$ models planetary dynamics ⇝ amount of incoming solar radiation)

1.23 FACTORS AND EXTENSIONS

Def: A mpt $(X, \mathcal{B}, \mu, T)$ is called a factor of a mpt $(Y, \mathcal{C}, \nu, S)$ if $\exists X' \subseteq X$, $\exists Y' \subseteq Y$, both of full measure, s.t.

- $T(X') \subseteq X'$, $S(Y') \subseteq Y'$
- there is a $\pi : Y' \to X'$ measurable and onto (surjective), with $\mu \circ \pi^{-1} = \nu$ and $\pi \circ S = T \circ \pi$ on Y'

π is called the factor map. $(Y, \mathcal{C}, \nu, S)$ is called an extension of $(X, \mathcal{B}, \mu, T)$.

Examples

1) Meas.-theoretically isomorphic systems are factors and extensions of each other
2) A skew-product $T : \Omega \times X \circlearrowleft$ is an extension of its base $\sigma : X \to X$, if μ is a probability measure.
3) Assume, we can "measure" the state $x \in X$ of some mpt only through the observables $f : X \to \mathbb{R}$ (measurable). The dynamical information we obtain is encoded in the sub-σ-alg

$$\mathcal{C} := \sigma\left(f \circ T^n \mid n \geq 0\right) = \sigma\left\{T^{-n} \circ f^{-1}(A) \mid A \in \mathcal{B}(\mathbb{R})\right\}$$

(E.g. $f = \chi_A$ for some $A \in \mathcal{B}$, and then we observe as trajectories sequences in $\{0,1\}^{\mathbb{N}}$)

It can happen that $(X, \mathcal{C}, \mu, T)$ is mixing, while $(X, \mathcal{B}, \mu, T)$ isn't.

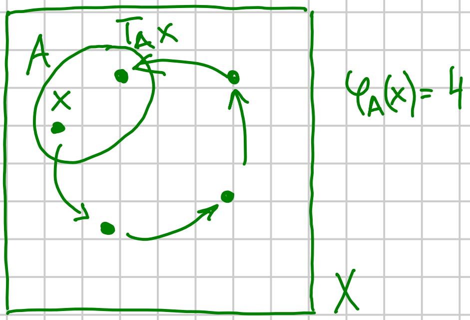
1.24 INDUCED TRANSFORMATION

Let $(X, \mathcal{B}, \mu, T)$ be a ppt and $A \in \mathcal{B}$ with $\mu(A) > 0$

PRT (§1.9) $\Longrightarrow$ for a.e. $x \in A$: $\exists n \geq 1$ s.t. $T^n(x) \in A$. So, let

$\varphi_A(x) := \min\{n \geq 1 \mid T^n(x) \in A\}$. (Note $\varphi_A < \infty$ a.e. on A)

Def: The induced transformation on A is $(A_0, \mathcal{B}(A), \mu_A, T_A)$, where



- $A_0 = \{x \in A \mid \varphi_A(x) < \infty\}$
- $\mathcal{B}(A) = \{E \cap A \mid A \in \mathcal{B}\}$
- $\mu_A(E) = \dfrac{\mu(E \cap A)}{\mu(A)}$
- $T_A : A_0 \to A_0$, $\quad T_A(x) = T^{\varphi_A(x)}(x)$

Thm: Let $(X,\mathcal{B},\mu,T)$ be a ppt, and $A\in\mathcal{B}$ with $\mu(A)>0$. Then

(a) $\mu_A\circ T_A^{-1}=\mu_A$ (invariance)

(b) T ergodic $\Rightarrow$ T_A ergodic (but T mixing $\not\Rightarrow$ T_A mixing)

(c) μ ergodic, then for $f\in L^1(X,\mu)$:

$$\int f\,d\mu = \int_A \sum_{k=0}^{\varphi_A-1} f\circ T^k\,d\mu$$

(Kac Formula)

Proof: (a) For $E\in\mathcal{B}$:

invariance of μ

$$\mu(E)=\mu(T^{-1}E)$$

$\mu(B)=\mu(B\cap A)+\mu(B\cap A^c)$

$$=\mu(\underbrace{T^{-1}E\cap A}_{=T_A^{-1}E\cap\{\varphi_A=1\}})+\mu(T^{-1}E\cap A^c)$$

inv. of μ

$$=\mu(T_A^{-1}E\cap\{\varphi_A=1\})+\underbrace{\mu\circ T^{-1}(T^{-1}E\cap A^c)}_{\|}$$

$$\mu(\underbrace{T^{-2}E\cap T^{-1}A^c\cap A}_{=T_A^{-1}E\cap\{\varphi_A=2\}})+\mu(T^{-2}E\cap T^{-1}A^c\cap A^c)$$

$$=\dots$$

$$=\sum_{k=1}^{n}\mu(T_A^{-1}E\cap\{\varphi_A=k\}) + \mu\Big(T^{-n}E\cap\bigcap_{j=0}^{n-1}T^{-j}A^c\Big)$$

for every $n\in\mathbb{N}$.

Since $\bigcup_{k=1}^{n} T_A^{-1}E\cap\{\varphi_A=k\}\uparrow T_A E$ $(n\to\infty)$, and the union is over disjoint sets, we have $\sum_{k=1}^{n}\dots\xrightarrow[n\to\infty]{}\mu(T_A^{-1}E)$, hence $\mu(E)\geq\mu(T_A^{-1}E)$. Now take $A\setminus E$ instead of E:

$$\mu(A\setminus E) \geq \mu\left(T_A^{-1}A \setminus T_A^{-1}E\right) \quad \Leftrightarrow$$

$$\mu(A) - \mu(E) \geq \mu(A) - \mu(T_A^{-1}E)$$

In summary: $\mu(E) = \mu(T_A^{-1}E) \overset{\cdot\frac{1}{\mu(A)}}{\Longrightarrow} \mu_A(E) = \mu_A(T_A^{-1}E)$

(b) and (c): see [Sa, Thm 1.6] ■

Note: With $f = \mathbb{1} = \chi_X$ in the Kac formula:

$$1 = \int_A \sum_{r=0}^{\varphi_A(x)-1} 1 \, d\mu(x) \Longrightarrow \underline{\underline{\int_A \varphi_A \, d\mu_A = \frac{1}{\mu(A)}}}$$

Interpretation: "The average return time to A is $\mu(A)^{-1}$; or: the frequency of visiting A is $\mu(A)$."

Compare with the Ergodic Hypothesis (§1.7)