Chapter 6

Qualitative statistical properties: general facts

From the previous chapter we learned that long time predictions may be impossible even for seemingly simple Dynamicl Systems. Yet, surprisingly, it is exactly such an unpredictability that makes statistical predictions possible. In this chapter we expalin how to make sense of sentences like: *such and such will happen with probability p*.

For simplicity we will maily consider dicrete Dynamcial Systems, eventhough we will briefly comment on flows.

6.1 Basic Definitions and examples

Definition 6.1.1 *By Dynamical System with discrete time we mean a triplet* (X, T, μ) *where* X *is a measurable space,*[1] μ *is a measure and* T *is a measurable map from* X *to itself that preserves the measure (i.e.,* $\mu(T^{-1}A) = \mu(A)$ *for each measurable set* $A \subset X$*).*

An equivalent characterization of invariant measure is $\mu(f \circ T) = \mu(f)$ for each $f \in L^1(X, \mu)$ since, for each measurable set A, $\mu(\chi_A \circ T) = \mu(\chi_{T^{-1}A}) = \mu(T^{-1}A)$, where χ_A is the characteristic function of the set A.

Remark 6.1.2 *In the following we will always assume* $\mu(X) < \infty$

[1] By measurable space we simply mean a set X together with a σ-algebra that defines the measurable sets.

(and quite often $\mu(X) = 1$, i.e. μ is a probability measure). Nevertheless, the reader should be aware that there exists a very rich theory pertaining to the case $\mu(X) = \infty$, see [Aar97].

Definition 6.1.3 *By Dynamical System with continuous time we mean a triplet (X, ϕ^t, μ) where X is a measurable space, μ is a measure and ϕ^t is a measurable group ($\phi^t(x)$ is a measurable function for each t, $\phi^t(x)$ is a measurable function of t for almost all $x \in X$; ϕ^0 =identity and $\phi^t \circ \phi^s = \phi^{t+s}$ for each t, $s \in \mathbb{R}$) or semigroup ($t \in \mathbb{R}^+$) from X to itself that preserves the measure (i.e., $\mu((\phi^t)^{-1}A) = \mu(A)$ for each measurable set $A \subset X$).*

The above definitions are very general, this reflects the wideness of the field of Dynamical Systems. In the present book we will be interested in much more specialized situations.

In particular, X will always be a topological compact space. The measures will alway belong to the class $\mathcal{M}^1(X)$ of Borel probability measures on X.[2] For future use, given a topological space X and a map T let us define $\mathcal{M}_T$ as the collection of all Borel measures that are T invariant.[3]

Often X will consist of finite unions of smooth manifolds (eventually with boundaries). Analogously, the dynamics (the map or the flow) will be smooth in the interior of X.

Let us see few examples to get a feeling of how a Dynamical System can look like.

6.1.1 Examples

Rotations

Let $\mathbb{T}$ be $\mathbb{R}$ mod 1. By this we mean $\mathbb{R}$ quotiented with respect to the equivalence relations $x \sim y$ if and only if $x - y \in \mathbb{Z}$. $\mathbb{T}$ can be though as the interval $[0, 1]$ with the points 0 and 1 identified. We put on it the topology induced by the topology of $\mathbb{R}$ via the defined equivalence relation. Such a topology is the usual one on $[0, 1]$, apart from the fact that each open set containing 0 must contain 1 as well. Clearly, from the topological

[2] Remember that a Borel measure is a measure defined on the Borel σ-algebra, that is the σ-algebra generated by the open sets.

[3] Obviously, for each $\mu \in \mathcal{M}_T$, (X, T, μ) is a Dynamical System.

point of view, $\mathbb{T}$ is a circle. We choose the Borel σ-algebra. By μ we choose the Lebesgue measure m, while $T : \mathbb{T} \to \mathbb{T}$ is defined by

$$Tx = x + \omega \mod 1,$$

for some $\omega \in \mathbb{R}$. In essence, T translates, or rotates, each point by the same quantity ω. It is easy to see that the measure μ is invariant (Problem 6.4).

Bernoulli shift

A Dynamical System needs not live on some differentiable manifold, more abstract possibilities are available.

Let $\mathbb{Z}_n = \{1, 2, ..., n\}$, then define the set of two sided (or one sided) sequences $\Sigma_n = \mathbb{Z}_n^{\mathbb{Z}}$ ($\Sigma_n^+ = \mathbb{Z}_n^{\mathbb{Z}_+}$). This means that the elements of Σ_n are sequences $\sigma = \{..., \sigma_{-1}, \sigma_0, \sigma_1, ...\}$ ($\sigma = \{\sigma_0, \sigma_1, ...\}$ in the one sided case) where $\sigma_i \in \mathbb{Z}_n$. To define the measure and the σ-algebra a bit of care is necessary. To start with, consider the *cylinder sets*, that is the sets of the form

$$A_i^j = \{\sigma \in \Sigma_n \mid \sigma_i = j\}.$$

Such sets will be our basic objects and can be used to generate the algebra $\mathcal{A}$ of the cylinder sets via unions and complements (or, equivalently, intersections and complements). We can then define a topology on Σ_n (the product topology, if $\{1, \dots, n\}$ is endowed by the discrete topology) by declaring the above algebra made of open sets and a basis for the topology. To define the σ-algebra we could take the minimal σ-algebra containing $\mathcal{A}$, yet this it is not a very constructive definition, neither a particular useful one, it is better to invoke the Carathèodory construction.

Let us start by defining a measure on $\mathbb{Z}_n$, that is n numbers $p_i > 0$ such that $\sum_{i=1}^n p_i = 1$. Then, for each $i \in \mathbb{Z}$ and $j \in \mathbb{Z}_n$,

$$\mu(A_i^j) = p_j.$$

Next, for each collection of sets $\{A_{i_l}^{j_l}\}_{l=1}^s$, with $i_l \neq i_k$ for each $l \neq k$, we define

$$\mu(A_{i_1}^{j_1} \cap A_{i_2}^{j_2} \cap ... \cap A_{i_s}^{j_s}) = \prod_{l=1}^s p_{j_l}.$$

We now know the measure of all finite intersection of the sets A_i^j. Obviously $\mu(A^c) := 1 - \mu(A)$ and the measure of the union of two sets A, B obviously must satisfy $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$. We have so defined μ on $\mathcal{A}$. It is easy to check that such a μ is σ-additive on $\mathcal{A}$; namely: if $\{A_i\} \subset \mathcal{A}$ are pairwise disjoint sets and $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$, then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$. The next step is to define an outer measure[4]

$$\mu^*(A) := \inf_{\substack{B \in \mathcal{A} \\ B \supset A}} \mu(B) \quad \forall A \subset \Sigma_n.$$

Finally, we can define the σ-algebra as the collection of all the sets that satisfy the *Carathèodory's criterion*, namely A is measurable (that is belongs to the σ-algebra) iff

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \quad \forall E \subset \Sigma_n.$$

The reader can check that the sets in $\mathcal{A}$ are indeed measurable.

The Carathèodory Theorem then asserts that the measurable sets form a σ-algebra and that on such a σ-algebra μ^* is numerably additive, thus we have our measure μ (simply the restriction of μ^* to the σ-algebra).[5] The σ-algebra so obtained is nothing else than the completion with respect to μ of the minimal σ-algebra containing $\mathcal{A}$ (all the sets with zero outer measure are measurable).

The map $T : \Sigma_n \to \Sigma_n$ (usually called *shift*) is defined by

$$(T\sigma)_i = \sigma_{i+1}.$$

We leave to the reader the task to show that the measure is invariant (see Problem 6.12).

To understand what's going on, let us consider the function $f : \Sigma \to \mathbb{Z}_n$ defined by $f(\sigma) = \sigma_0$. If we consider T^t, $t \in \mathbb{N}$, as the time evolution and f as an observation, then $f(T^t\sigma) = \sigma_t$. This can be interpreted as the observation of some phenomenon at various times. If we do not know anything concerning the state of the system, then the probability to see

[4] An outer measure has the following properties: i) $\mu^*(\emptyset) = 0$; ii) $\mu^*(A) \leq \mu^*(B)$ if $A \subset B$; iii)$\mu^*(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$. Note that μ^* need not be additive on all sets.

[5] See [LL01] if you want a quick look at the details of the above Theorem or consult [Roy88] if you want a more in depth immersion in measure theory. If you think that the above construction is too cumbersome see Problem 6.14.

the value j at the time t is simply p_j. If $n = 2$ and $p_1 = p_2 = \frac{1}{2}$, it could very well be that we are observing the successive outcomes of tossing a fair coin where 1 means head and 2 tail (or vice versa); if $n = 6$ it could be the outcome of throwing a dice and so on.

Dilation

Again $X = \mathbb{T}$ and the measure is Lebesgue. T is defined by

$$Tx = 2x \mod 1.$$

This map it is not invertible (similarly to the one sided shift). Note that, in general, $\mu(TA) \neq \mu(A)$ (e.g., $A = [0, \frac{1}{2}]$).

Toral automorphism (Arnold cat)

This is an automorphism of the torus and gets its name by a picture draw by Arnold [AA68]. The space X is the two dimensional torus $\mathbb{T}^2$. The measure is again Lebesgue measure and the map is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \mod 1 := L\begin{pmatrix} x \\ y \end{pmatrix} \mod 1.$$

Since the entries of L are integers numbers it is clear that T is well defined on the torus; in fact, it is a linear toral automorphism. The invariance of the measure follows from $\det L = 1$.

Hamiltonian Systems

Up to now we have seen only examples with discrete time. Typical examples of Dynamical Systems with continuous time are the solutions of an ODE or a PDE. Let us consider the case of an Hamiltonian system. The simplest case is when $X = \mathbb{R}^{2n}$, the σ-algebra is the Borel one and the measure μ is the Lebesgue measure m. The dynamics is defined by a smooth function $H : X \to \mathbb{R}$ via the equations

$$\frac{dx}{dt} = J\text{grad}H(x)$$

where $\text{grad}(H)_i = (\nabla H)_i = \frac{\partial H}{\partial x_i}$ and J is the block matrix

$$J = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix}.$$

The fact that m is invariant with respect to the Hamiltonian flow is due to the Liouville Theorem (see [Arn99] or Problem 5.7).

Such a dynamical system has a natural decomposition. Since H is an integral of the motion, for each $h \in \mathbb{R}$ we can consider $X_h = \{x \in X \mid H(x) = h\}$. If $X_h \neq \emptyset$, then it will typically consist of a smooth manifold,[6] let us restrict ourselves to this case. Let σ be the surface measure on X_h, then $\mu_h = \frac{\sigma}{\|\mathrm{grad} H\|}$ is an invariant measure on X_h and (X_h, ϕ_t, μ_h) is a Dynamical System (see Problem 6.6).

Geodesic flow

Along the same lines any geodesic flow on a compact Riemannian manifold naturally defines a dynamical system.

6.2 Return maps and Poincaré sections

Normally in Dynamical Systems there is a lot of emphasis on the discrete case. One reason is that there is a general device that allows to reduce the study of many properties of a continuous time Dynamical System to the study of an appropriate discrete time Dynamical System: Poincaré sections (we have already seen an instance of this in the introduction). Here we want to make few comments on this precious tool that we will largely employ in the study of billiards.

Let us consider a smooth Dynamical System (X, ϕ^t, μ) (that is a Dynamical Systems in continuous time where X is a smooth manifold and ϕ^t is a smooth flow). Then we can define the vector field $V(x) := \frac{d\phi^t(x)}{dt}\big|_{t=0}$.[7]

Consider a smooth compact submanifold (possibly with boundaries) Σ of codimension one such that $\mathcal{T}_x\Sigma$ (the tangent space of Σ at the point x) is transversal to $V(x)$.[8] We can then define the *return time* $\tau_\Sigma : \Sigma \to \mathbb{R}^+ \cup \{\infty\}$ by

$$\tau_\Sigma = \inf\{t \in \mathbb{R}^+ \backslash \{0\} \mid \phi^t(x) \in \Sigma\},$$

[6] By the implicit function theorem this is locally the case if $\nabla H \neq 0$.

[7] Very often it is the other way around: the vector field is given first and then the flow–as we saw in the introduction.

[8] That is $\mathcal{T}_x\Sigma \oplus V(x)$ form the full tangent space at x.

where the inf is taken to be ∞ if the set is empty. Next we define the *return map* $T_\Sigma : D(T) \subset \Sigma \to \Sigma$, where $D(T) = \{x \in \Sigma | \tau_\Sigma(x) < \infty\}$, by

$$T_\Sigma(x) = \phi^{\tau_\Sigma(x)}(x).$$

It is easy to check that there exists $c > 0$ such that $\tau_\Sigma \geq c$ (Problem 6.9).

To define the measure, the natural idea is to project the invariant measure along the flow direction: for all measurable sets $A \subset \Sigma$, define[9]

$$\nu_\Sigma(A) := \lim_{\delta \to 0} \frac{1}{\delta} \mu(\phi^{[0,\delta]}(A)). \tag{6.2.1}$$

See Problem **??** for the existence of the above limit; see Problem 6.9 for the proof that τ_Σ is finite almost everywhere and Problem 6.10 for the proof that $(\Sigma, T_\Sigma, \nu_\Sigma)$ is a dynamical system. The reader is invited to meditate on the relation between this Dynamical System and the original one.

6.3 Suspension flows

A natural question is if it is possible to construct a flow with a given Poincaré section, the answer is that there are infinitely many flows with a given section. Let us construct some of them. Given a dynamical system (Σ, T, ν) consider $\tilde{X} := \Sigma \times R^+$. Define the flow $\phi_t((x, s)) = (x, s + t)$. We then define in $\tilde{X}$ the equivalence relation $(x, t) \sim (y, s)$ iff $s = t + n$ and $y = T^n x$ or $t = s + n$ and $x = T^n y$ for some $n \in \mathbb{N}$. A moment of reflection shows that the set X of equivalence classes is nothing else than the set $\Sigma \times [0, 1]$ with the points $(x, 1)$ and $(Tx, 0)$ identified. Clearly the flow is naturally quotiented over the equivalence classes and yields a quotient flow on X, such a flow is called a *suspension flow*.

A more general construction can by obtained by applying a time change to the above example. Alternatively, one can can choose any smooth function $\tau : \Sigma \to \mathbb{R}^+$, that will be called a *ceiling function* and consider the set $X_\tau = \{(x, t) \in \Sigma \times \mathbb{R}^+ \mid t \in [0, \tau(x)]\}$ with the points $(x, \tau(x))$ and $(Tx, 0)$ identified. A moment of reflection should

[9] We use the notation: $\phi^I(A) := \cup_{t \in I} \phi^t(A)$ for each $I \subset \mathbb{R}$.

show that the topology of X_τ does not depend on τ and is then the same than the suspension defined above. The flow is again defined by $\phi_t(x,s) = (x, s+t)$ for $t \le \tau(x) - s$. Such flows are called *special flows.*

6.4 Invariant measures

A very natural question is: given a space X and a map T does there always exists an invariant measure μ? A non exhaustive, but quite general, answer exists: Krylov-Bogoluvov Theorem.

First of all we need a useful characterization of invariance.

Lemma 6.4.1 *Given a compact metric space X and map T continuous apart from a compact set K,*[10] *a Borel measure μ, such that $\mu(K) = 0$, is invariant if and only if $\mu(f \circ T) = \mu(f)$ for each $f \in \mathcal{C}^0(X)$.*

PROOF. To prove that the invariance of the measure implies the invariance for continuous functions is obvious since each such function can be approximate uniformly by simple functions–that is, sum of characteristic functions of measurable sets–for which the invariance it is immediate.[11] The converse implication is not so obvious.

The first thing to remember is that the Borel measures, on a compact metric space, are regular [RS80]. This means that for each measurable set A the following holds[12]

$$\mu(A) = \inf_{\substack{G \supset A \\ G = \mathring{G}}} \mu(G) = \sup_{\substack{C \subset A \\ C = \overline{C}}} \mu(C). \tag{6.4.2}$$

Next, remember that for each closed set A and open set $G \supset A$, there exists $f \in \mathcal{C}^0(X)$ such that $f(X) \subset [0,1]$, $f|_{G^c} = 0$ and $f|_A = 1$ (this is Urysohn Lemma for Normal spaces [Roy88]). Hence, setting $B_A := \{f \in \mathcal{C}^{(0)}(X) \mid f \ge \chi_A\}$,

$$\mu(A) \le \inf_{f \in B_A} \mu(f) \le \inf_{\substack{G \supset A \\ G = \mathring{G}}} \mu(G) = \mu(A). \tag{6.4.3}$$

[10] This means that, if $C \subset X$ is closed, then $T^{-1}C \cup K$ is closed as well.

[11] This is essentially the definition of integral.

[12] This is rather clear if one thinks of the Carathéodory construction starting from the open sets.

Accordingly, for each A closed, we have

$$\mu(T^{-1}A) \leq \inf_{f \in B_A} \mu(f \circ T) = \inf_{f \in B_A} \mu(f) = \mu(A).$$

In addition, using again the regularity of the measure, for each A Borel holds[13]

$$\begin{aligned}\mu(T^{-1}A) &= \inf_{\substack{U \supset K \\ U = \mathring{U}}} \mu(T^{-1}A \backslash U) \leq \inf_{\substack{U \supset K \\ U = \mathring{U}}} \sup_{\substack{C \subset T^{-1}A \backslash U \\ C = \overline{C}}} \mu(T^{-1}(TC)) \\ &\leq \inf_{\substack{U \supset K \\ U = \mathring{U}}} \sup_{\substack{C \subset A \backslash TU \\ C = \overline{C}}} \mu(T^{-1}C) \leq \sup_{\substack{C \subset A \\ C = \overline{C}}} \mu(T^{-1}C) = \sup_{\substack{C \subset A \\ C = \overline{C}}} \mu(C) = \mu(A).\end{aligned}$$

Applying the same argument to the complement A^c of A it follow that it must be $\mu(T^{-1}A) = \mu(A)$ for each Borel set. □

Proposition 6.4.2 (Krylov–Bogoluvov) *If X is a metric compact space and $T : X \to X$ is continuous, then there exists at least one invariant (Borel) measure.*

Proof. Consider any Borel probability measure ν and define the following sequence of measures $\{\nu_n\}_{n \in \mathbb{N}}$:[14] for each Borel set A

$$\nu_n(A) = \nu(T^{-n}A).$$

The reader can easily see that $\nu_n \in \mathcal{M}^1(X)$, the sets of the probability measures. Indeed, since $T^{-1}X = X$, $\nu_n(X) = 1$ for each $n \in \mathbb{N}$. Next, define

$$\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} \nu_i.$$

Again $\mu_n(X) = 1$, so the sequence $\{\mu_i\}_{i=1}^{\infty}$ is contained in a weakly compact set (the unit ball) and therefore admits a weakly convergent

[13] Note that, by hypothesis, if C is compact and $C \cap K = \emptyset$, then TC is compact.

[14] Intuitively, if we chose a point $x \in X$ at random, according to the measure ν and we ask what is the probability that $T^n x \in A$, this is exactly $\nu(T^{-n}A)$. Hence, our procedure to produce the point $T^n x$ is equivalent to picking a point at random according to the evolved measure ν_n.

subsequence $\{\mu_{n_i}\}_{i=1}^{\infty}$; let μ be the weak limit.[15] We claim that μ is T invariant. Since μ is a Borel measure it suffices to verify that for each $f \in \mathcal{C}^0(X)$ holds $\mu(f \circ T) = \mu(f)$ (see Lemma 6.4.1). Let f be a continuous function, then by the weak convergence we have[16]

$$\mu(f \circ T) = \lim_{j\to\infty} \frac{1}{n_j} \sum_{i=0}^{n_j-1} \nu_i(f \circ T) = \lim_{j\to\infty} \frac{1}{n_j} \sum_{i=0}^{n_j-1} \nu(f \circ T^{i+1})$$
$$= \lim_{j\to\infty} \frac{1}{n_j} \left\{ \sum_{i=0}^{n_j-1} \nu_i(f) + \nu(f \circ T^{n_j}) - \nu(f) \right\} = \mu(f).$$

□

The reason why the above theorem is not completely satisfactory is that it is not constructive and, in particular, does not provide any information on the nature of the invariant measure. On the contrary, in many instances the interest is focused not just on any Borel measure but on special classes of measures, for example measures connected to the Lebesgue measure which, in some sense, can be thought as reasonably physical measures (if such measures exists).

In the following examples we will see two main techniques to study such problems: on the one hand it is possible to try to construct explicitly the measure and study its properties in the given situations (expanding maps, strange attractors, solenoid, horseshoe); on the other hand one can try to *conjugate*[17] the given problem with another, better

[15]This depends on the Riesz-Markov Representation Theorem [RS80] that states that $\mathcal{M}(X)$ is exactly the dual of the Banach space $\mathcal{C}^0(X)$. Since the weak convergence of measures in this case correspond exactly to the weak-* topology [RS80], the result follows from the Banach-Alaoglu theorem stating that the unit ball of the dual of a Banach space is compact in the weak-* topology. But see 1.6.17 if you want a more elementary proof.

[16]Note that it is essential that we can check invariance only on continuous functions: if we would have to check it with respect to all bounded measurable functions we would need that μ_n converges in a stronger sense (*strong convergence*) and this may not be true. Note as well that this is the only point where the continuity of T is used: to insure that $f \circ T$ is continuous and hence that $\mu_{n_j}(f \circ T) \to \mu(f \circ T)$.

[17]See Definition 6.8.2 for a precise definition and Problem 6.37 and 6.38 for some insight.

understood, one (logistic map, circle maps). In view of the second possibility the last example is very important (Markov measures). Such an example gives just a hint to the possibility to construct a multitude of invariant measures for the shift which, as we will see briefly, is a standard system to which many other can be conjugated.

6.4.1 Examples

Contracting maps

Let $X \subset \mathbb{R}^n$ be compact and connected, $T : X \to X$ differentiable with $\|DT\| \leq \lambda^{-1} < 1$ and $T0 = 0 \in X$. In this case 0 is the unique fixed point and the delta function at zero is the only invariant measure.[18]

Expanding maps

The simplest possible case is $X = \mathbb{T}$, $T \in \mathcal{C}^2(\mathbb{T})$ with $|DT| \geq \lambda > 1$, (see Figure 6.1 for a pictorial example).[19]

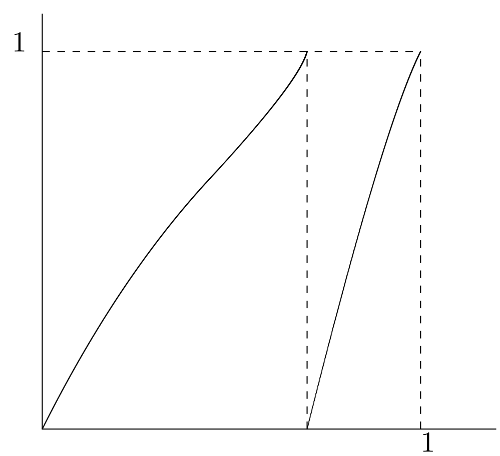

Figure 6.1: Graph of an expanding map on $\mathbb{T}$

[18] The reader will hopefully excuse this physicist language, naturally we mean that the invariant measure is defined by $\delta_0(f) = f(0)$. The property that there exists only one invariant measure is called *unique ergodicity*, we will see more of it in the sequel, e.g. see example 6.5.1.

[19] Note that this generalizes Examples 6.1.1.

We would like to have an invariant measure absolutely continuous with respect to Lebesgue. Any such measure μ has, by definition, the Radon-Nikodym derivative $h = \frac{d\mu}{dm} \in L^1(\mathbb{T}, m)$, [Roy88]. In Proposition 6.4.2 we saw how a measure evolves by defining the operator

$$T_*\mu(f) = \mu(f \circ T) \tag{6.4.4}$$

for each $f \in \mathcal{C}^0$ and $\mu \in \mathcal{M}(X)$ (see also footnote 15 at page 126). If we want to study a smaller class of measures we must first check that T_* leaves such a class invariant. Indeed, if μ is absolutely continuous with respect to Lebesgue then $T_*\mu$ has the same property. Moreover, if $h = \frac{d\mu}{dm}$ and $h_1 = \frac{dT_*\mu}{dm}$ then (Problem 6.15)

$$h_1(x) = \mathcal{L}h(x) := \sum_{y \in T^{-1}(x)} |D_yT|^{-1}h(y).$$

The operator $\mathcal{L} : L^1(\mathbb{T}, m) \to L^1(\mathbb{T}, m)$ is called *Transfer operator* or Ruelle-Perron-Frobenius operator, and has an extremely important rôle in the study of the statistical properties of the system. Notice that $\|\mathcal{L}h\|_1 \leq \|h\|_1$.[20] The key property of $\mathcal{L}$, in this context, is given by the following inequality (this type of inequality is commonly called of Lasota-York type) (Problem 6.16)

$$\left|\frac{d}{dx}\mathcal{L}h(x)\right| \leq \lambda^{-1}|\mathcal{L}h'(x)| + C|\mathcal{L}h(x)| \tag{6.4.5}$$

where $C = \frac{\|D^2T\|_\infty}{\|DT\|_\infty^2}$.

The above inequality implies $\|(\mathcal{L}h)'\|_1 \leq \lambda^{-1}\|h'\|_1 + C\|h\|_1$. Iterating such a relation yields

$$\|(\mathcal{L}^n h)'\|_1 \leq \frac{C}{1-\lambda^{-1}}\|h\|_1 + \|h'\|_1,$$

for all $n \in \mathbb{N}$. This, in turn, implies that the $\sup_{n\in\mathbb{N}} \|\mathcal{L}^n h\|_\infty < \infty$. Consequently, the sequence $h_n := \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{L}^i h$ is compact in L^1 (this is a consequence of standard embedding theorems[21] [LL01] but see Problem

[20] Here $\|f\|_1 := \int |h(x)|dx$ is the standard norm in L^1.

[21] Indeed the space $\mathcal{C}^1$ closed with respect to the norm $\|f\| = \|f\|_1 + \|f'\|_1$ is a well known Banach space: the Sobolev space $W^{1,1}$.

6.17 for an elementary proof). In analogy with Lemma 6.4.2, we have that there exists $h_* \in L^1$ such that $\mathcal{L}h_* = h_*$. Thus $d\mu := h_* dm$ is an invariant measure of the type we are looking for.

In fact, it is possible to obtain some more information on such measure. Equation 6.4.5 implies that $\mathcal{L}$ is a well defined operator also when restricted to $\mathcal{C}^0$ or $\mathcal{C}^1$. Moreover, for each $h \in \mathcal{C}^0$ and $n \in \mathbb{N}$,

$$|\mathcal{L}^n h|_\infty \leq |\mathcal{L}^n 1|_\infty |h|_\infty \leq |h|_\infty (\|\mathcal{L}^n 1\|_1 + \|(\mathcal{L}^n 1)'\|_1) \leq |h|_\infty \frac{C+1}{1-\lambda^{-1}} =: C_1 |h|_\infty.$$

Using the above equation and iterating (6.4.5) yields, for each $h \in \mathcal{C}^1$ and $n \in \mathbb{N}$,

$$|(\mathcal{L}^n h)'|_\infty \leq \lambda^{-n} C_1 |h'|_\infty + C_1^2 |h|_\infty.$$

In other words we have a Lasota-Yorke type inequality for $\mathcal{L}$ acting on $\mathcal{C}^0$, $\mathcal{C}^1$ instead of $L^1, W^{1,1}$. In particular note that one can apply the above inequalities to the average $h_n := \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{L}^i h$, when $h \in \mathcal{C}^1$. Then the compactness follows by Ascoli-Arzelá Theorem and it follows that the invariant density is continuous (in fact, Lipschitz as already argued in the Perron-Frobenius Theorem).

Logistic maps

Consider $X = [0,1]$ and

$$T(x) = 4x(1-x).$$

This map is not an everywhere expanding map ($D_{\frac{1}{2}}T = 0$), yet it can be conjugate with one, [UvN47].

To see this consider the continuous change of variables $\Psi : [0,1] \to [0,1]$ defined by

$$\Psi(x) = \frac{2}{\pi} \arcsin \sqrt{x},$$

thus $\Psi^{-1}(x) = \left(\sin \frac{\pi}{2} x\right)^2$. Accordingly,

$$\begin{aligned} \tilde{T}(x) \quad &:= \Psi \circ T \circ \Psi^{-1}(x) = \Psi(4 \sin^2 \tfrac{\pi}{2} x \cos^2 \tfrac{\pi}{2} x) \\ &= \Psi([\sin \pi x]^2) = \tfrac{2}{\pi} \arcsin[\sin \pi x] \end{aligned}$$

which yields[22]

$$\tilde{T}(x) = \begin{cases} 2x & \text{for } x \in [0, \frac{1}{2}] \\ 2 - 2x & \text{for } x \in [\frac{1}{2}, 1]. \end{cases}$$

The map $\tilde{T}$ is called *tent map* for its characteristic shape, see figure

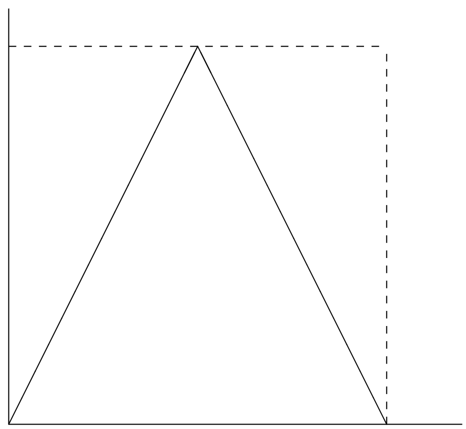

Figure 6.2: Graph of tent map

6.2. What is more interesting is that the Lebesgue measure is invariant for $\tilde{T}$, as the reader can easily check. This means that, if we define $\mu(f) := m(f \circ \Psi^{-1})$, it holds true

$$\mu(f \circ T) = m(f \circ T \circ \Psi^{-1}) = m(f \circ \Psi^{-1} \circ \tilde{T}) = m(f \circ \Psi^{-1}) = \mu(f).$$

Hence, $([0,1], T, \mu)$ is a Dynamical System. In addition, a trivial computation shows

$$\mu(dx) = \frac{1}{\pi\sqrt{x(1-x)}} dx,$$

thus μ is absolutely continuous with respect to Lebesgue.

Circle maps

A circle map is an order preserving continuous map of the circle. A simple way to describe it is to start by considering its lift. Let $\hat{T} : \mathbb{R} \to \mathbb{R}$, such

[22] Remember that the domain of arcsin is $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $\sin \pi x = \sin \pi(1-x)$.

that $\hat{T}(0) \in [0, 1]$, $\hat{T}(x+1) = \hat{T}(x)+1$ ad it is monotone increasing. The circle map is then defined as $T(x) = \hat{T}(x) \mod 1$. Circle maps have a very rich theory that we do not intend to develop here, we confine ourselves to some facts (see [HK95] for a detailed discussion of the properties below). The first fact is that the *rotation number*

$$\rho(T) = \lim_{n\to\infty} \frac{1}{n}\hat{T}^n(x).$$

is well defined and does not depend on x.

We have already seen a concrete example of circle maps: the rotation R_ω by ω. Clearly $\rho(R_\omega) = \omega$. It is fairly easy to see that if $\rho(T) \in \mathbb{Q}$ then the map has a periodic orbit. We are more interested in the case in which the rotation number is irrational. In this case, with the extra assumption that T is twice differentiable (actually a bit less is needed) the Denjoy theorem holds stating that there exists a continuous invertible function h such that $R_{\rho(T)} \circ h = h \circ T$, that is T is *topologically conjugated* to a rigid rotation. Since we know that the Lebesgue measure is invariant for the rotations, we can obtain an invariant measure for T by pushing the Lebesgue measure by h, namely define

$$\mu(f) = m(f \circ h^{-1}).$$

The natural question if the measure μ is absolutely continuous with respect to Lebesgue is rather subtle and depends, once again, on KAM theory. In essence the answer is positive only if T has more regularity and the rotation number is not very well approximated by rational numbers (in some sense it is 'very irrational').

Strange Attractors

We have seen the case in which all the trajectories are attracted by a point. The reader can probably imagine a case in which the attractor is a curve or some other simple set. Yet, it has been a fairly recent discovery that an attractor may have a very complex (strange) structure. The following is probably the simplest example. Let $X = Q = [0, 1]^2$ and

$$T(x, y) = \begin{cases} (2x,\, \frac{1}{8}y + \frac{1}{4}) & \text{if } x \in [0,\, 1/2] \\ (2x-1,\, \frac{1}{8}y + \frac{3}{4}) & \text{if } x \in]1/2,\, 1]. \end{cases}$$

We have a map of the square that stretches in one direction by a factor 2 and contract in the other by a factor 8.

Note that T it is not continuous with respect to the normal topology, so Proposition 6.4.2 cannot be applied directly. This problem can be solved in at least two ways: one is to *code* the system and we will discuss it later (see Examples 6.8.1), the other is to study more precisely what happens iterating a measure in special cases.

In our situation, since $T^n Q$ consists of a multitude of thinner and thinner strips, it is clear that there can be no invariant measure absolutely continuous with respect to Lebesgue.[23] Yet, it is very natural to ask what happens if we iterate the Lebesgue measure by the operator T_*. It is easy to see that $T_* m$ is still absolutely continuous with respect to Lebesgue. In fact, T_* maps absolutely continuous measures into absolutely continuous measures. Once we note this, it is very tempting to define the transfer operator. An easy computation yields

$$\mathcal{L}h(x) = \chi_{TQ}(x) \sum_{y \in T^{-1}(x)} |\det(D_y T)|^{-1} h(y) = 4\chi_{TQ}(x) h(T^{-1}(x)).$$

Since the map expands in the unstable direction, it is quite natural to investigate, in analogy with the expanding case, the *unstable derivative* D^u, that is the derivative in the x direction, of the iterate of the density.

$$\|D^u \mathcal{L}h\|_1 \leq \frac{1}{2} \|D^u h\|_1 \quad \forall h \in \mathcal{C}^1(Q) \tag{6.4.6}$$

To see the consequences of the above estimate, consider $f \in \mathcal{C}^{(1)}(Q)$ with $f(0, y) = f(1, y) = 0$ for each $y \in [0, 1]$, then if μ is a measure obtained by the measure hdm ($h \in \mathcal{C}^1$) with the procedure of Proposition 6.4.2,[24]

[23] In fact, if μ is an invariant measure, $T_*\mu = \mu$, it follows

$$\mu(\chi_{T^n Q}) = T_*^n \mu(\chi_{T^n Q}) = \mu(\chi_Q) = 1,$$

so μ must be supported on $\Lambda = \cap_{n=0}^{\infty} T^n Q$.

[24] As we noted in the proof of Proposition 6.4.2, the only part that uses the continuity of T is the proof of the invariance. Thus, in general we can construct a measure by the averaging procedure but its invariance is not automatic.

we have

$$\mu(D^u f) = \lim_{j\to\infty} \frac{1}{n_j} \sum_{i=0}^{n_j-1} (T_*)^i m(hD^u f) = \lim_{j\to\infty} \frac{1}{n_j} \sum_{i=0}^{n_j-1} m(\mathcal{L}^i h D^u f)$$
$$= -\lim_{j\to\infty} \frac{1}{n_j} \sum_{i=0}^{n_j-1} m(f D^u \mathcal{L}^i h)$$

where we have integrated by part. Remembering (6.4.6) we have

$$\mu(D^u f) = 0,$$

for all $f \in \mathcal{C}^{(1)}_{\text{per}}(Q) = \{f \in \mathcal{C}^{(1)}(Q) \mid f(0,y) = f(1,y)\}$. The enlargement of the class of functions is due to the obvious fact that, if $f \in \mathcal{C}^{(1)}_{\text{per}}(Q)$, then $\tilde{f}(x,y) = f(x,y) - f(0,y)$ is zero on the vertical (stable) boundary and $D^u \tilde{f} = D^u f$.

This means that the measure μ, when restricted to the horizontal direction, is μ-a.e. constant (see Problem 6.32). Such a strong result is clearly a consequence of the fact that the map is essentially linear, one can easily imagine a non linear case (think of dilations and expanding maps) and in that case the same argument would lead to conclude that the measure, when restricted to unstable manifolds, is absolutely continuous with respect to the restriction of Lebesgue (these type of measures are commonly called SRB from Sinai, Ruelle and Bowen).

We can now prove that indeed the measure μ is invariant. The discontinuity line of T is $\{x = \frac{1}{2}\}$. Points close to $\{x = \frac{1}{2}\}$ are mapped close to the boundary of Q, so if $f(0,y) = f(1,y) = 0$, then $f \circ T$ is continuous. Hence, the argument of Proposition 6.4.2 proves that $\mu(f \circ T) = \mu(f)$ for all f that vanish at the stable boundary. Yet, the characterization of μ proves that $\mu(\{(x,y) \in Q \mid x \in \{0,1\}\}) = 0$, thus we can obtain $\mu(f \circ T) = \mu(f)$ for all continuous functions via the Lebesgue dominated convergence theorem and the invariance follows by Lemma 6.4.1.

Horseshoe

This very famous example consists of a map of the square $Q = [0,1]^2$, the map is obtained by stretching the square in the horizontal direction, bending it in the shape of an horseshoe and then superimposing it to the original square in such a way that the intersection consists of two horizontal strips.[25] Such a description is just topological, to make things clearer let us consider a very special case:

$$T(x,\,y) = \begin{cases} (5x \mod 1,\, \frac{1}{4}y) & \text{if } x \in [1/5,\, 2/5] \\ (5x \mod 1,\, \frac{1}{4}y + \frac{3}{4}) & \text{if } x \in [3/5,\, 4/5]. \end{cases}$$

Note that T is not explicitly defined for $x \in [0,\, 1/5[\cup[\frac{2}{3},\, \frac{3}{5}[\cup]4/5,\, 1]$ since for this values the horseshoe falls outside Q, so its actual shape is irrelevant. Since the map from Q to Q is not defined on the full square, we can have a Dynamical System only with respect to a measure for which the domain of definition of T, and all of its powers, has measure one. We will start by constructing such a measure.

The first step is to notice that the set

$$\Lambda = \cap_{n\in\mathbb{Z}} T^n Q \tag{6.4.7}$$

of the points which trajectories are always in Q is $\neq \emptyset$. Second, note that $\Lambda = T\Lambda = T^{-1}\Lambda$, such an invariant set is called *hyperbolic set* as we will see in ???. We would like to construct an invariant measure on Λ. Since Λ is a compact set and T is continuous on it we know that there exist invariant measures; yet, in analogy with the previous examples, we would like to construct one *coming from Lebesgue*.

As already mentioned we must start by constructing a measure on $\Lambda_- = \cap_{n\in\mathbb{N}\cup\{0\}} T^{-n}Q$ since $T^k\Lambda_- \subset \Lambda_-$. To do so it is quite natural to construct a measure by *subtracting* the mass that leaks out of Q. namely, define the operator $\tilde{T} : \mathcal{M}(X) \to \mathcal{M}(X)$ by

$$\tilde{T}\mu(A) := \mu(TA \cap Q).$$

Again we consider the evolution of measures of the type $d\mu = hdm$. For each continuous f with $\text{supp}(f) \subset Q$ holds

$$\tilde{T}\mu(f) = \mu(f \circ T^{-1}\chi_Q) = \int_{T^{-1}Q} fh \circ T|\det DT|dm.$$

We can thus define the operator $\mathcal{L}$ that evolves the densities:

$$\mathcal{L}h(x) = \frac{5}{4}\chi_{T^{-1}Q\cap Q}(x)h(Tx).$$

[25] We have already seen something very similar in the introduction.

Clearly $\tilde{T}\mu(f) = m(f\mathcal{L}h)$.

Note that $\tilde{T}m(1) = \frac{1}{2}$, thus $\tilde{T}$ does not map probability measures into probability measures; this is clearly due to the mass leaking out of Q. Calling D^s (stable derivative) the derivative in the y direction, follows easily

$$\|D^s\mathcal{L}h\|_1 \leq \frac{1}{4}\|D^s h\|_1$$

for each h differentiable in the stable direction.

On the other hand, if $\|D^s h\|_1 \leq c$ and $\Delta = [0, 1/4] \cup [3/4, 1]$,

$$\begin{aligned}
|\tilde{T}\mu(1)| &= \int_{Q\cap TQ} h = \int_{\Delta} dy \int_0^1 dx h(x,y) \\
&= \int_{\Delta} dy \int_0^1 dx \int_0^1 d\xi h(x,\xi) + \mathcal{O}(\|D^s h\|_1) \\
&= |\Delta| \|h\|_1 + \mathcal{O}(\|D^s h\|_1) = \frac{1}{2}\mu(1) + \mathcal{O}(\|D^s h\|_1).
\end{aligned}$$

It is then natural to define $\hat{\mathcal{L}}h := 2\mathcal{L}h$ and $\hat{T} = 2\tilde{T}$. Thus $\|D^s\hat{\mathcal{L}}h\|_1 \leq \frac{1}{2}\|D^s h\|_1$. This means that $\{\frac{1}{n}\sum_{i=0}^{n-1}\hat{T}^i\mu\}$ are probability measures. Accordingly, there exists an accumulation point μ_* and $\mu_*(D^s f) = 0$ for each f periodic in the y direction. By the same type of arguments used in the previous examples, this means that μ_* is constant in the y direction, it is supported on Λ_- by construction and $\tilde{T}\mu_* = \frac{1}{2}\mu_*$ ($conformal\ invariance$) : just the measure we where looking for.

We can now conclude the argument by evolving the measure as usual:

$$T_*\mu_*(f) = \mu_*(f \circ T)$$

for all continuous f with the support in Q. Now the standard argument applies. In such a way we have obtained the invariant measure supported on Λ.

Markov Measures

Let us consider the shift (Σ_n^+, T). We would like to construct other invariant measures bedside Bernoulli. As we have seen it suffices to specify the measure on the algebra of the cylinders. Let us define

$$A(m; k_1, \dots, k_l) = \{\sigma \in \Sigma_n^+ \mid \sigma_{i+m} = k_i \ \forall\, i \in \{1, \dots, l\}\};$$

this are a basis for the algebra of the cylinders.

For each $n \times n$ matrix P, $P_{ij} \geq 0$, $\sum_j P_{ij} = 1$ by the Perron-Frobenius theorem (see Secion (A.2.2)) there exists $\{p_i\}$ such that $pP = p$. Let us define

$$\mu(A(m; k_1, \ldots, k_l)) = p_{k_1} P_{k_1 k_2} P_{k_2 k_3} \ldots P_{k_{l-1} k_l}.$$

The reader can easily verify that μ is invariant over the algebra $\mathcal{A}$ and thus extends to an invariant measure. This is called Markov because it is nothing else than a Markov chain together with its stationary measure.[26]

These last examples (strange attractor, solenoid, horseshoe) show only a very dim glimpse of a much more general and extremely rich theory (the study of SRB measures) while the last (Markov measures) points toward another extremely rich theory: Gibbs (or equilibrium) measures. Although this it is not the focus here, we will see a bit more of this in the future.

One of the main objectives in dynamical systems is the study of the long time behavior (that is the study of the trajectories $T^n x$ for large n). There are two main cases in which it is possible to study, in some detail, such a long time behavior. The case in which the motion is rather regular[27] or close to it (the main examples of this possibility are given by the so called KAM [Arn92] theory and by situations in which the motions is attracted by a simple set); and the case in which the motion is very irregular.[28] This last case may seem surprising since the irregularity of the motion should make its study very difficult. The reason why such systems can be studied is, as usual, because we ask the right questions,[29] that is we ask questions not concerning the fine details of the motion but only concerning its statistical or qualitative properties.

The first example of such properties is the study of the invariant sets.

[26] The probabilistic interpretation is that the probability of seeing the state k at time one, given that we saw the state l at time zero, is given by P_{lk}. So the process has a bit of memory: it remembers its state one time step before. Of course it is possible to consider processes that have a longer–possibly infinite–memory. Proceeding in this direction one would define the so called *Gibbs measures*.

[27] Typically, quasi periodic motion, remember the small oscillation in the pendulum.

[28] Remember the example in the introduction.

[29] Of course, the "right questions" are the ones that can be answered.

6.5 Ergodicity

Definition 6.5.1 *A measurable set A is invariant for T if $T^{-1}A \subset A$.*

A dynamical system (X, T, μ) is ergodic if each invariant set has measure zero or one.

The definition for continuous dynamical systems being exactly the same.

Note that if A is invariant then $\mu(A \backslash T^{-1}A) = \mu(A) - \mu(T^{-1}A) = 0$, moreover $\Lambda = \cap_{n=0}^{\infty} T^{-n}A \subset A$ is invariant as well. In addition, by definition, $\Lambda = T\Lambda$, which implies $\Lambda = T^{-1}\Lambda$ and $\mu(A\backslash\Lambda) = 0$. This means that, if A is invariant, then it always contains a set Λ invariant in the stronger (maybe more natural) sense that $T\Lambda = T^{-1}\Lambda = \Lambda$. Moreover, Λ is of full measure in A. Our definition of invariance is motivated by its greater flexibility and the fact that, from a measure theoretical point of view, zero measure sets can be discarded.

In essence, if a system is ergodic then most trajectories explore all the available space. In fact, for any A of positive measure, define $A_b = \cup_{n\in\mathbb{N}\cup\{0\}} T^{-n}A$ (this are the points that eventually end up in A), since $A_b \supset A$, $\mu(A_b) > 0$. Since $T^{-1}A_b \subset A_b$, by ergodicity follows $\mu(A_b) = 1$. Thus, the points that never enter in A (that is, the points in A_b^c) have zero measure. Actually, if the system has more structure (topology) more is true (see Problem 6.21).

The reader should be aware that there are many equivalent definitions of ergodicity, see Problems 6.25, 6.27, 6.28 and Theorem 6.6.6 for some possibilities.

6.5.1 Examples

Rotations

The ergodicity of a rotation depends on ω. If $\omega \in \mathbb{Q}$ then the system is not ergodic. In fact, let $\omega = \frac{p}{q}$ $(p, q \in \mathbb{N})$, then, for each $x \in \mathbb{T}$ $T^q x = x + p \mod 1 = x$, so T^q is just the identity. An alternative way of saying this is to notice that all the points have a periodic trajectory of period q. It is then easy to exhibit an invariant set with measure strictly larger than 0 but strictly less than 1. Consider $[0, \varepsilon]$, then $A = \cup_{i=1}^{q-1} T^{-i}[0, \varepsilon]$ is an invariant set; clearly $\varepsilon \leq \mu(A) \leq q\varepsilon$, so it suffices to choose $\varepsilon < q^{-1}$.

The case $\omega \notin \mathbb{Q}$ is much more interesting. First of all, for each point $x \in \mathbb{T}$ we have that the closure of the set $\{T^n x\}_{i=0}^{\infty}$ is equal to $\mathbb{T}$, which is to say that the orbits are dense.[30] The proof is based on the fact that there cannot be any periodic orbit. To see this suppose that $x \in \mathbb{T}$ has a periodic orbit, that is there exists $q \in \mathbb{N}$ such that $T^q x = x$. As a consequence there must exist $p \in \mathbb{Z}$ such that $x + p = x + q\omega$ or $\omega \in \mathbb{Q}$ contrary to the hypothesis. Hence, the set $\{T^k 0\}_{k=0}^{\infty}$ must contain infinitely many points and, by compactness, must contain a convergent subsequence k_i. Hence, for each $\varepsilon > 0$, there exists $m > n \in \mathbb{N}$:

$$|T^m 0 - T^n 0| < \varepsilon.$$

Since T preserves the distances, calling $q = m - n$, holds

$$|T^q 0| < \varepsilon.$$

Accordingly, the trajectory of $T^{jq}0$ is a translation by a quantity less than ε, therefore it will get closer than ε to each point in $\mathbb{T}$ (i.e., the orbit is dense). Again by the conservation of the distance, since zero has a dense orbit the same will hold for every other point.

Intuitively, the fact that the orbits are dense implies that there cannot be a non trivial invariant set, henceforth the system is ergodic. Yet, the proof it is not trivial since it is based on the existence of Lebesgue density points [Roy88] (see Problem 6.40). It is a fact from general measure theory that each measurable set $A \subset \mathbb{R}$ of positive Lebesgue measure contains, at least, one point $\bar{x}$ such that for each $\varepsilon \in (0, 1)$ there exists $\delta > 0$:

$$\frac{m(A \cap [\bar{x} - \delta, \bar{x} + \delta])}{2\delta} > 1 - \varepsilon.$$

Hence, given an invariant set A of positive measure and $\varepsilon > 0$, first choose δ such that the interval $I := [\bar{x} - \delta, \bar{x} + \delta]$ has the property $m(I \cap A) > (1 - \varepsilon)m(I)$. Second, we know already that there exists $q, M \in \mathbb{N}$ such that $\{T^{-kq}x\}_{k=1}^{M}$ divides $[0, 1]$ into intervals of length less that $\frac{\varepsilon}{2}\delta$. Hence, given any point $x \in \mathbb{T}$ choose $k \in \mathbb{N}$ such that $m(T^{-kq}I \cap [x - \delta, x + \delta]) > m(I)(1 - \varepsilon)$ so,

$$\begin{aligned} m(A \cap [x - \delta, x + \delta]) &\geq m(A \cap T^{-kq}I) - m(I)\varepsilon \\ &\geq m(A \cap I) - m(I)\varepsilon \geq (1 - 2\varepsilon)2\delta. \end{aligned}$$

[30] A system with a dense orbit called *Topologically Transitive*.

Thus, A has density everywhere larger than $1-2\varepsilon$, which implies $\mu(A)=1$ since ε is arbitrary.

The above proof of ergodicity it is not so trivial but it has a definite dynamical flavor (in the sense that it is obtained by studying the evolution of the system). Its structure allows generalizations to contexts whit a less rich algebraic structure. Nevertheless, we must notice that, by taking advantage of the algebraic structure (or rather the group structure) of $\mathbb{T}$, a much simpler and powerful proof is available.

Let $\nu \in \mathcal{M}_T^1$, then define

$$F_n = \int_{\mathbb{T}} e^{2\pi i n x} \nu(dx), \quad n \in \mathbb{N}.$$

A simple computation, using the invariance of ν, yields

$$F_n = e^{2\pi i n \omega} F_n$$

and, if ω is irrational, this implies $F_n = 0$ for all $n \neq 0$, while $F_0 = 1$. Next, consider $f \in \mathcal{C}^{(2)}(\mathbb{T}^1)$ (so that we are sure that the Fourier series converges uniformly, see Problem 6.31), then

$$\nu(f) = \sum_{n=0}^{\infty} \nu(f_n e^{2\pi i n \cdot}) = \sum_{n=0}^{\infty} f_n F_n = f_0 = \int_{\mathbb{T}} f(x) dx.$$

Hence m is the unique invariant measure (unique ergodicity). This is clearly much stronger than ergodicity (see Problem 6.25)

Expanding maps

Next, we prove that any smooth invariant map has a unique invariant measure absolutely continuos with respect to Lebesgue and hence it is ergodic with respect to such a measure. Let $h \in L^1$ be the density of an invariant measure and A, of positive measure, an invariant set. For each $\varepsilon > 0$ there exists $f_\varepsilon \in \mathcal{C}^1$ such that $\|f_\varepsilon - \mathbb{1}_A\|_1 \leq \varepsilon$. Calling $f_{\varepsilon,n} = \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{L}^i f_\varepsilon$ and noting that, by invariance, $\varphi_n := \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{L}^i \mathbb{1}_A = \mathbb{1}_A \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{L}^i 1$, we have, by taking subsequeces, that f_n converges in $\mathcal{C}^0$ to some invariant density $\bar{f}_\varepsilon$ while φ_n converges to $\mathbb{1}_A h$, where h is the invariant density to which converges $\frac{1}{n}\sum_{i=0}^{n-1} \mathcal{L}^i 1$ (or rather the chosen subsequence). On the other hand $\|\bar{f}_\varepsilon - \mathbb{1}_A h\|_1 \leq \varepsilon$. Since the $\bar{f}_\varepsilon$ are all uniformly Lipschitz, hence equicontinuous, (see the end of Example 6.4.1, Expanding maps) by Ascoli-Arzelá we can extract a converging subquence. This means that $\mathbb{1}_A$ is the uniform limit of continuos functions, hence it is continuos hence A is either empty of everything, thus the map is ergodic. The uniqueness of the invariant measure follows by similar arguments.

Baker

This transformation gets its name from the activity of bread making, it bears some resemblance with the horseshoe. The space X is the square $[0,\,1]^2$, μ is again Lebesgue, and T is a transformation obtained by squashing down the square into the rectangle $[0,\,2]\times[0,\,\frac{1}{2}]$ and then cutting the piece $[1,\,2]\times[0,\,\frac{1}{2}]$ and putting it on top of the other one. In formulas

$$T(x,\,y)=\begin{cases}(2x,\,\frac{1}{2}y)\mod 1 & \text{if } x\in[0,\,\frac{1}{2})\\ (2x,\,\frac{1}{2}(y+1))\mod 1 & \text{if } x\in[\frac{1}{2},\,1].\end{cases}$$

This transformation is ergodic as well, in fact much more. We will discuss it later.

Translations ($\mathbb{T}^1$)

Let us consider the flow $(\mathbb{T}^1,\phi_t,m)$ where $\phi_t(x)=x+\omega t \mod 1$, for some $\omega\in\mathbb{R}\setminus\{0\}$. This is just a translation on the unit circle. The proof of ergodicity is trivial and it is left to the reader.

We conclude the chapter with a theorem very helpful to establish the ergodicity of a flow.

Theorem 6.5.2 *Consider a flow (X,ϕ_t,μ) and a Poincarè section Σ such that the set $\{x\in X\mid \cup_{t\in\mathbb{R}}\phi_t(x)\cap\Sigma=\emptyset\}$ has zero measure. Then the ergodicity of the flow (X,ϕ_t,μ) is equivalent to the ergodicity of the section $(\Sigma,T_\Sigma,\mu_\Sigma)$.*

The proof, being straightforward, is left to the reader.

6.5.2 Examples

Translations **($\mathbb{T}^2$)**

Let us consider the flow $(\mathbb{T}^2, \phi_t, m)$ where $\phi_t(x) = x + \omega t \mod 1$, for some $\omega \in \mathbb{R}^2 \setminus \{0\}$. This is a translation on the two dimensional torus. To investigate we will use Theorem 6.5.2. Consider the set $\Sigma := \{(x, y) \in \mathbb{T}^2 \mid x = 0\}$, this is clearly a Poincaré section, unless $\omega_1 = 0$ (in which case one can choose the section $y = 0$). Obviously Σ is a circle and the Poincaré map is given by

$$T(y) = y + \frac{\omega_2}{\omega_1} \mod 1.$$

The ergodicity of the flow is then reduced to the ergodicity of a circle rotation, thus the flow is ergodic only if ω_1 and ω_2 have an irrational ratio.

The properties of the invariant sets of a dynamical systems have very important reflections on the statistics of the system, in particular on its time averages. Before making this precise (see Theorem 6.6.6) we state few very general and far reaching results.

6.6 Some basic Theorems

Theorem 6.6.1 *(Birkhoff) Let (X, T, μ) be a dynamical system, then for each $f \in L^1(X, \mu)$*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)$$

exists for almost every point $x \in X$. In addition, setting

$$f^+(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x),$$

holds

$$\int_X f^+ d\mu = \int_X f d\mu.$$

Proof

Since the task at hand is mainly didactic, we will consider explicitly only the case of positive bounded functions, the completion of the proof is left to the reader.

Let $f \in L^\infty(X, d\mu)$, $f \geq 0$, and

$$S_n(x) \equiv \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x).$$

For each $x \in X$, there exists

$$\begin{aligned}\overline{f}^+(x) &= \limsup_{n\to\infty} S_n(x) \\ \underline{f}^+(x) &= \liminf_{n\to\infty} S_n(x).\end{aligned}$$

The first remark is that both $\overline{f}^+$ and $\underline{f}^+$ are invariant functions. In fact,

$$S_n(Tx) = S_n(x) + \frac{1}{n}f(T^n x) - \frac{1}{n}f(x)$$

so, tacking the limit the result follows.[31]

Next, for each $n \in \mathbb{N}$ and k, $j \in \mathbb{Z}$ we define

$$D_{n,l,j} = \left\{ x \in X \;\middle|\; \overline{f}^+(x) \in \left[\frac{l}{n}, \frac{l+1}{n}\right) ;\; \underline{f}^+(x) \in \left[\frac{j}{n}, \frac{j+1}{n}\right) \right\},$$

by the invariance of the functions follows the invariance of the sets $D_{n,l,j}$. Also, by the boundedness, follows that for each n exists n_0 such as

$$\bigcup_{j,l \in \{-n_0, \dots, n_0\}} D_{n,l,j} = X.$$

The key observation is the following.

Lemma 6.6.2 *For each $n \in \mathbb{N}$ and l, $j \in \mathbb{Z}$, setting $A = D_{n,l,j}$, holds*

$$\begin{aligned}\frac{l+1}{n}\mu(A) &< \int_A f d\mu + \frac{3}{n}\mu(A) \\ \frac{j}{n}\mu(A) &> \int_A f d\mu - \frac{3}{n}\mu(A)\end{aligned}$$

[31] Here we have used the boundedness, this is not necessary. If $f \in L^1(X, d\mu)$ and positive, then $S_n(Tx) \geq S_n(x) - f(x)$, so $\underline{\overline{f}}^+(Tx) \geq \underline{\overline{f}}^+(x)$ and it is and easy exercise to check that any such function must be invariant.

From the Lemma follows

$$0 \leq \int_X (\overline{f}^+ - \underline{f}^+) d\mu = \sum_{l,\, j=-n_0}^{n_0} \int_{D_{n,l,j}} (\overline{f}^+ - \underline{f}^+) d\mu$$
$$\leq \sum_{l,\, j=-n_0}^{n_0} \left[\frac{l+1}{n} - \frac{j}{n} \right] \mu(D_{n,l,j}) < \frac{6}{n} \sum_{l,\, j=-n_0}^{n_0} \mu(D_{n,l,j}) = \frac{6}{n}.$$

Since n is arbitrary we have

$$\int_X (\overline{f}^+ - \underline{f}^+) d\mu = 0$$

which implies $\overline{f}^+ = \underline{f}^+$ almost everywhere (since $\overline{f}^+ \geq \underline{f}^+$ by definition) proving that the limit exists. Analogously, we can prove

$$\int_X (f - f^+) d\mu = 0.$$

Proof of the Lemma 6.6.2 We will prove only the first inequality, the second being proven in exactly the same way.

For each $x \in A$ we will call $k(x)$ the first $m \in \mathbb{N}$ such that

$$S_m(x) > \frac{l-1}{n},$$

by construction $k(x)$ must be finite for each $x \in A$. Hence, setting $X_k = \{x \in A \mid k(x) = k\}$, $\cup_k X_k = A$, and for each $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\mu\left(\bigcup_{k=1}^{N} X_k\right) \geq \mu(A)(1-\varepsilon).$$

Let us call

$$Y = A \backslash \bigcup_{k=1}^{N} X_k.$$

Then $\mu(Y) \leq \mu(A)\varepsilon$, also set $L = \sup_{x \in A} |f(x)|$. The basic idea is to follow, for each point $x \in A$, the trajectory $\{T^i x\}_{i=0}^{M}$, where $M > N$ will be chosen sufficiently large. If the point would never visit the set Y, we could group the sum $S_M(x)$ in pieces all, in average, larger than

$\frac{l-1}{n}$, so the same would hold for $S_M(x)$. The difficulties come from the visits to the set Y.

For each $n \in \{0, ..., M\}$ define

$$\widetilde{f}_n(x) = \begin{cases} f(T^n x) & \text{if } T^n x \notin Y \\ \dfrac{l}{n} & \text{if } T^n x \in Y \end{cases}$$

and

$$\widetilde{S}_M(x) = \frac{1}{M} \sum_{n=0}^{M-1} \widetilde{f}_n(x).$$

By definition $y \in Y$ implies $y \notin X_1$, i.e. $f(y) \leq \frac{l-1}{n}$. Accordingly, $\widetilde{f}(x) \geq f(T^n x)$ for each $x \in A$. Note that for each n we change the function $f \circ T^n$ only at some points belonging to the set Y and $\frac{l}{n}$ can be taken less or equal than L (otherwise $\mu(A) = 0$), consequently

$$\int_A f d\mu = \int_A S_M d\mu \geq \int_A \widetilde{S}_M d\mu - L\mu(Y) \geq \int_A \widetilde{S}_M d\mu - L\mu(A)\varepsilon.$$

We are left with the problem of computing the sum. As already mentioned the strategy consists in dividing the points according to their trajectory with respect to the sets X_n. To be more precise, let $x \in A$, then by definition it must belong to some X_n or to Y. We set $k_1(x)$ equal to j is $x \in X_j$ and $k_1(x) = 1$ if $x \in Y$. Next, $k_2(x)$ will have value j if $T^{k_1(x)} x \in X_j$ or value 1 if $T^{k_1(x)} \in Y$. If $k_1(x) + k_2(x) < M$, then we go on and define similarly $k_3(x)$. In this way, to each $x \in A$ we can associate a number $m(x) \in \{1, ..., M\}$ and indices $\{k_i(x)\}_{i=1}^{m(x)}$, $k_i(x) \in \{1, ..., N\}$, such that $M - N \leq \sum_{i=1}^{m(x)-1} k_i(x) < M$, $\sum_{i=1}^{m(x)} k_i(x) \geq M$. Let us call $K_p(x) = \sum_{j=1}^{p} k_j(x)$. Using such a division of the orbit in segments of length $k_i(x)$ we can easily estimate

$$\begin{aligned}
\widetilde{S}_M(x) =& \frac{1}{M} \left\{ \sum_{i=1}^{m(x)-1} k_i(x) \left[\frac{1}{k_i(x)} \sum_{j=K_{i-1}(x)}^{K_i(x)-1} \widetilde{f}_j(x) \right] + \sum_{i=K_{m(x)-1}(x)}^{M-1} \widetilde{f}(T^i x) \right\} \\
\geq& \frac{1}{M} \sum_{i=1}^{m(x)-1} k_i(x) \frac{l-1}{n} \geq \frac{M-N}{M} \frac{l-1}{n}.
\end{aligned}$$

Putting together the above inequalities we get

$$\int_A f d\mu \geq \left\{ \frac{(M-N)(l-1)}{Mn} - L\varepsilon \right\} \mu(A)$$
$$\geq \frac{l+1}{n}\mu(A) - \left\{ \frac{2}{n} + \frac{N(l-1)}{Mn} + L\varepsilon \right\} \mu(A).$$

which, by choosing first ε sufficiently small and, after, M sufficiently large, concludes the proof. □

To prove the result for all function in $L^1(X, \mu)$ it is convenient to deal at first only with positive functions (which suffice since any function is the difference of two positive functions) and then use the usual trick to cut off a function (that is, given f define f_L by $f_L(x) = f(x)$ if $f(x) \leq L$, and $f_L(x) = L$ otherwise) and then remove the cut off. The reader can try it as an exercise. □

Birkhoff theorem has some interesting consequences.

Corollary 6.6.3 *For each $f \in L^1(X, \mu)$ the following holds*

1. $f_+ \in L^1(X, \mu)$;
2. $f_+(Tx) = f_+(x)$ *almost surely.*

The proof is left to the reader as an easy exercise (see Problem 6.18).

Another interesting fact, that starts to show some connections between averages and invariant sets, emerges by considering a measurable set A and its characteristic function χ_A. A little thought shows that the ergodic average $\chi_A^+(x)$ is simply the average frequency of visit of the set A by the trajectory $\{T^n x\}$ (Problem 6.28).

Birkhoff theorem implies also convergence in L^1 and L^2 (see also Problem 6.26). Yet, it is interesting to note that convergence in L^2 can be proven in a much more direct way.

Theorem 6.6.4 (Von Neumann) *Let (X, T, μ) be a Dynamical System, then for each $f \in L^2(X, \mu)$ the ergodic average converges in $L^2(X, \mu)$.*

Proof. We have already seen that it can be useful to lift the dynamics at the level of the algebra of function or at the level of measures.

This game assumes different guises according to how one plays it, here is another very interesting version.

Let us define $U : L^2(X,\mu) \to L^2(X,\mu)$ as

$$Uf := f \circ T.$$

Then, by the invariance of the measure, it follows $\|Uf\|_2 = \|f\|_2$, so U is an L^2 contraction (actually, and L^2-isometry). If T is invertible, the same argument applied to the inverse shows that U is indeed unitary, otherwise we must content ourselves with

$$\|U^*f\|_2^2 = \langle UU^*f, f\rangle \le \|UU^*f\|_2\|f\|_2 = \|U^*f\|_2\|f\|_2,$$

that is $\|U^*\|_2 \le 1$ (also U^* is and L^2 contraction).

Next, consider $V_1 = \{f \in L^2 \mid Uf = f\}$ and $V_2 = \text{Rank}(\mathbb{1} - U)$. First of all, note that if $f \in V_1$, then

$$\|U^*f - f\|_2^2 = \|U^*f\|_2^2 - \langle f, U^*f\rangle - \langle U^*f, f\rangle + \|f\|_2^2 \le 0.$$

Thus, $f \in V_1^* := \{f \in L^2 \mid U^*f = f\}$. The same argument applied to $f \in V_1^*$ shows that $V_1 = V_1^*$. To continue, consider $f \in V_1$ and $h \in L^2$, then

$$\langle f, h - Uh\rangle = \langle f - U^*f, h\rangle = 0.$$

This implies that $V_1^\perp = \overline{V_2}$, hence $V_1 \oplus \overline{V_2} = L^2$. Finally, if $g \in V_2$, then there exists $h \in L^2$ such that $g = h - Uh$ and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{\infty} U^i g = \lim_{n\to\infty}\frac{1}{n}(h - U^n h) = 0.$$

On the other hand if $f \in V_1$ then $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{\infty} U^i f = f$. The only function on which we do not still have control are the g belonging to the closure of V_2 but not in V_2. In such a case there exists $\{g_k\} \subset V_2$ with $\lim_{k\to\infty} g_k = g$. Thus,

$$\|\frac{1}{n}\sum_{i=0}^{\infty} U^i g\|_2 \le \|\frac{1}{n}\sum_{i=0}^{\infty} U^i g_k\|_2 + \|g - g_k\|_2 \le \|\frac{1}{n}\sum_{i=0}^{\infty} U^i g_k\|_2 + \frac{\varepsilon}{2},$$

provided we choose k large enough. Then, by choosing n sufficiently large we obtain

$$\|\frac{1}{n}\sum_{i=0}^{\infty} U^i g\|_2 \le \varepsilon.$$

We have just proven that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}U^i = P$$

where P is the orthogonal projection on V_1. □

Another very general result, of a somewhat disturbing nature, is Poincaré return theorem.

Theorem 6.6.5 (Poincaré) *Given a dynamical systems $(X,\,T,\,\mu)$ and a measurable set A, with $\mu(A) > 0$, there exists infinitely many $n \in \mathbb{N}$ such that*

$$\mu(T^{-n}A \cap A) \neq 0.$$

The proof is rather simple (by contradiction) and the reader can certainly find it out by herself (see Problem 6.19).[32]

Let us go back to the relation between ergodicity and averages. From an intuitive point of view a function from X to $\mathbb{R}$ can be thought as an "observable," since to each configuration it associates a value that can represent some relevant property of the configuration (the property that we observe). So, if we observe the system for a long time via the function f, what we see should be well represented by the function f^+. Furthermore, notice that there is a simple relations between invariant functions and invariant sets. More precisely, if a measurable set A is invariant, then its characteristic function χ_A is a measurable invariant function; if f is an invariant function then for each measurable set $I \in \mathbb{R}$ the set $f^{-1}(I)$ is a measurable invariant set (if the implications of the above discussions are not clear to you, see Problem 6.27).

As a byproduct of the previous discussion it follows that if a system is ergodic then for each function $f \in L^1(X,\,\mu)$ the function f_+ is

[32] An unsettling aspect of the theorem is due to the following possibility. Consider a room full of air, the motion of the molecules can be thought to happen accordingly to Newton equations, i.e. it is an Hamiltonian systems, hence a dynamical system to which Poincaré theorem applies. Let A be the set of configurations in which all the air is in the left side of the room. Since we ignore, in general, the past history of the room, it could very well be that at some point in the past the systems was in a configuration belonging to A–maybe some silly experiment was performed. So there is a positive probability for the system to return in the same state. Therefore the disturbing possibility of sudden death by decompression.

almost everywhere constant and equal to $\int_X f$. We have just proven an interesting characterization of the ergodic systems:

Theorem 6.6.6 *A Dynamical System (X, T, μ) is ergodic if and only if for each $f \in L^1(X, \mu)$ the ergodic average f^+ is constant; in fact, $f^+ = \mu(f)$ a.e..*

In other words, if we observe the time average of some observable for a sufficiently long time then we obtain a value close to its space average. The previous observation is very important especially because the space average of a function does not depend on the dynamics. This is exactly what we where mentioning previously: the fact that the dynamics is sufficiently 'complex' allows us to ignore it completely, provided we are interested only in knowing some average behavior. The relevance of ergodic theory for physical systems is largely connected to this fact.

6.7 Mixing

We have argued the importance of ergodicity, yet from a physical point of view ergodicity may be relevant only if it takes places at a sufficiently fast rate (i.e., if the time average converges to the space average on a physically meaningful time scale). This has prompted the study of stronger statistical properties of which we will give a brief, and by no mean complete, account in the following.

Definition 6.7.1 *A Dynamical System (X, T, μ) is called mixing if for every pairs of measurable sets A, B we have*

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

Obviously, if a system is mixing, then it is ergodic. In fact, if A is an invariant set for T, then $T^{-n}A \subset A$, so, calling A^c the complement of A, we have

$$\mu(A)\mu(A^c) = \lim_{n\to\infty} \mu(T^{-n}A \cap A^c) = 0,$$

and the measure of A is either one or zero.

An equivalent characterization of mixing is the following:

Proposition 6.7.2 *A Dynamical System* (X, T, μ) *is mixing if and only if*

$$\lim_{n\to\infty}\int_X f\circ T^n g d\mu = \int_X f d\mu \int_X g d\mu$$

for every $f,\, g \in L^2(X,\, \mu)$ *or for every* $f \in L^\infty(X, \mu)$ *and* $g \in L^1(X, \mu)$.[33]

The proof is rather straightforward and it is left as an exercise to the reader (see Problem 6.29) together with the proof of the next statement.

Proposition 6.7.3 *A Dynamical System* $(X,\, T,\, \mu)$*, with* X *a compact metric space,* T *continuous and* μ *Borel, is mixing if and only if for each probability measure* λ *absolutely continuous with respect to* μ

$$\lim_{n\to\infty}\lambda(f\circ T^n) = \mu(f)$$

for each $f \in \mathcal{C}^0(\mathbb{T}^2)$.

This last characterization is interesting from a mathematical point of view. Define, as usual, the evolution of a measure via the equation

$$(T_*\lambda)(f) \equiv \lambda(f\circ T)$$

for each continuous function f. If for each measure, absolutely continuous with respect to the invariant one, the evolved measure converges weakly to the invariant measure, then the system is mixing (and thus the evolved measures converge strongly). This has also a very important physical meaning: if the initial configuration is known only in probability, the probability distribution is absolutely continuous with respect to the invariant measure, and the system is mixing, then, after some time, the configurations are distributed according to the invariant measure. Again the details of the evolution are not important to describe relevant properties of the system.

6.7.1 Examples

Rotations

We have seen that the translations by an irrational angle are ergodic. They are not mixing. The reader can easily see why.

[33]The quantity $\int_X f\circ Tg - \int_X f \int_X g$ is called "correlation," and its tending to zero–which takes places always in mixing systems–it is called "decay of correlation."

Bernoulli shift

The key observation is that, given a measurable set A, for each $\varepsilon > 0$ there exists a set $A_\varepsilon \in \mathcal{A}$, thus depending only on a finite subset of indices,[34] with the property[35]

$$\mu(A_\varepsilon \backslash A) \leq \varepsilon.$$

Then, given A, B measurable, and for each $\varepsilon > 0$, let A_ε, B_ε be such an approximation, and I_A, I_B the defining sets of indices, then

$$\left|\mu(T^{-m}A \cap B) - \mu(A)\mu(B)\right| \leq 4\varepsilon + \left|\mu(T^{-m}A_\varepsilon \cap B_\varepsilon) - \mu(A_\varepsilon)\mu(B_\varepsilon)\right|.$$

If we choose m so large that $(I_A + m) \cap I_B = \emptyset$, then by the definition of Bernoulli measure we have

$$\mu(T^{-m}A_\varepsilon \cap B_\varepsilon) = \mu(T^{-m}A_\varepsilon)\mu(B_\varepsilon) = \mu(A_\varepsilon)\mu(B_\varepsilon),$$

which proves

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

Dilation

This system is mixing. In fact, let $f,\, g \in \mathcal{C}^1(\mathbb{T})$, then we can represent them via their Fourier series $f(x) = \sum_{k\in\mathbb{Z}} e^{2\pi i k x} f_k$, $f_{-k} = \overline{f}_k$. It is well known that $\sum_{k\in\mathbb{Z}} |f_k| < \infty$ and $|f_k| \leq \frac{c}{|k|}$, for some constant c depending on f. Therefore,

$$f(T^n x) = \sum_{k\in\mathbb{Z}} e^{2\pi i 2^n k x} f_k,$$

which implies that the only Fourier coefficients of $f \circ T^n$ different from zero are the $\{2^n k\}_{k\in\mathbb{Z}}$. Hence,

$$\left|\int_{\mathbb{T}} f \circ T^n g - \int_{\mathbb{T}} f \int_{\mathbb{T}} g\right| = \left|\sum_{k\in\mathbb{Z}} f_k g_{2^n k} - f_0 g_0\right| \leq c 2^{-n} \sum_{k\in\mathbb{Z}} |f_k|.$$

The previous inequalities imply the exponential decay of correlations for each smooth function. The proof is concluded by a standard approximation

[34] Remember, this means that there exists a finite set $I \subset \mathbb{Z}$ such that it is possible to decide if $\sigma \in \Sigma_n$ belongs or not to A_ε only by looking at $\{\sigma_i\}_{i\in I}$.

[35] This follows from our construction of the σ-algebra and by the definition of outer measure, see Examples 6.1.1–Bernoulli shift.

argument: given $f,\, g \in L^2(X,\, d\mu)$, for each $\varepsilon > 0$ exists $f_\varepsilon,\, g_\varepsilon \in \mathcal{C}^1(X)$: $\|f - f_\varepsilon\|_2 < \varepsilon$ and $\|g - g_\varepsilon\|_2 < \varepsilon$. Thus,

$$\left|\int_{\mathbb{T}} f \circ T^n g - \int_{\mathbb{T}} f \int_{\mathbb{T}} g\right| \leq \left|\int_{\mathbb{T}} f_\varepsilon \circ T^n g_\varepsilon - \int_{\mathbb{T}} f_\varepsilon \int_{\mathbb{T}} g_\varepsilon\right| + 2(\|f\|_2 + \|g\|_2)\varepsilon,$$

which yields the result by choosing first ε small and then n sufficiently large.

6.8 Stronger statistical properties

One very fruitful idea in the realm of measurable dynamical systems is the idea of *entropy*. In some sense the entropy measure the complexity of the motions from a measure theoretical point of view.

To define it one starts by considering a partition of the space into measurable sets $\xi := \{A_1, \dots A_n\}$ and defines[36]

$$H_\mu(\xi) - \sum_i \mu(A_i) \log \mu(A_i).$$

Given two partitions $\xi = \{A_i\}, \eta = \{B_j\}$ we define $\xi \vee \eta := \{A_i \cap B_j\}$. Let then be

$$\xi_{-n}^T := \xi \vee T^{-1}(\xi) \vee \cdots \vee T^{-n+1}(\xi).$$

It is then possible to prove that the sequence $H_\mu(\xi_{-n}^T)$ is sub-additive, hence the limit

$$h_\mu(T, \xi) := \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_{-n}^T$$

exists.

Definition 6.8.1 *The entropy of T with respect to μ is defined as*

$$h_\mu(T) := \sup\{h_\mu(T, \xi) \mid H(\xi) < \infty\}$$

Clearly if a system has positive metric entropy this means that the motion has a high complexity and it is very far from regular. One of the main property of entropy is that it is a metric invariant, that is

[36] The case of a countable partition, or even an uncountable partition, can be handled and it is very relevant, but outside the aims of this book, see [Roh67] for a complete treatment of the subject.

if two systems are metrically conjugate (see the following), then they have the same metric entropy.

Even more extreme form statistical behaviors are possible, to present them we need to introduce the idea of equivalent systems. This is done via the concept of conjugation that we have already seen informally in Example 6.4.1 (logistic map, circle map).

Definition 6.8.2 *Two Dynamical Systems* (X_1, T_1, μ_1), (X_2, T_2, μ_2) *are (measurably) conjugate if there exists a measurable map* $\phi : X_1 \to X_2$ *almost everywhere invertible*[37] *such that* $\mu_1(A) = \mu(\phi(A))$ *and* $T_2 \circ \phi = \phi \circ T_1$.

Clearly, the conjugation is an equivalence relation. Its relevance for the present discussion is that conjugate systems have the same ergodic properties (Problem 6.38).[38]

We can now introduce the most extreme form of stochasticity.

Definition 6.8.3 *A dynamical system* (X, T, μ) *is called Bernoulli if there exists a Bernoulli shift* (M, ν, σ) *and a measurable isomorphism* $\phi : X \to M$ *(i.e., a measurable map one one and onto apart from a set of zero measure and with measurable inverse) such that, for each* $A \in X$,

$$\nu(\phi(A)) = \mu(A)$$

and

$$T = \phi^{-1} \circ \sigma \circ \phi.$$

That is a system is Bernoulli if it is isomorphic to a Bernoulli shift. Since we have seen that Bernoulli systems are very stochastic (remind that they can be seen as describing a random event like coin tossing) this is certainly a very strong condition on the systems. In particular it is immediate to see that Bernoulli systems are mixing (Problem 6.38).

[37] This means that there exists a measurable function $\phi^{-1} : X_2 \to X_1$ such that $\phi \circ \phi^{-1} = \text{id}$ μ_2-a.e. and $\phi^{-1} \circ \phi = \text{id}$ μ_1-a.e..

[38] Of course the reader can easily imagine other forms of conjugacy, e.g. topological or differential conjugation.

6.8.1 Examples

Dilation

We will show that such a system is indeed Bernoulli. The map ϕ is obtained by dividing $[0, 1)$ in $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1)$. Then, given $x \in \mathbb{T}$, we define $\phi : \mathbb{T} \to \Sigma_2^+$ by

$$\phi(x)_i = \begin{cases} 1 & \text{if } T^i x \in [0, \dfrac{1}{2}) \\ 2 & \text{if } T^i x \in [\dfrac{1}{2}, 1) \end{cases}$$

the reader can check that the map is measurable and that it satisfy the required properties. Note that the above shows that the Bernoulli measure with $p_1 = p_2 = \frac{1}{2}$ is nothing else than Lebesgue measure viewed on the numbers written in basis two. This may explain why we had to be so careful in the construction of the Bernoulli measure.

Baker

Let us define ϕ^{-1}; for each $\sigma \in \Sigma_2$

$$x = \sum_{i=0}^{\infty} \frac{\sigma_{-i}}{2^{i+1}},$$
$$y = \sum_{i=1}^{\infty} \frac{\sigma_i}{2^i}.$$

Again the rest is left to the reader.

Forced Pendulum

In the introduction we have seen that there exists a square Q with stable and unstable sides such that, calling T the map introduced by the flow at a proper time, $TQ \cap Q \supset Q_0^u \cup Q_1^u$. Where Q_i^u are rectangles that go from one stable side of Q to the other and, in analogy, $T^{-1}Q \cap Q \supset Q_0^s \cup Q_1^s$.

We can use this fact to code the dynamics similarly to what we have done for the Backer map. Namely, given the set $\Lambda = \bigcap_{n \in \mathbb{Z}} T^n Q$ (this set it is non empty–see Example 6.4.1–Horseshoe) and $\phi : \Lambda \to \Sigma_2$ define by

$$[\phi(x)]_k = \begin{cases} i \in \{0,1\} & \text{if } k \geq 0 \text{ and } T^k x \in Q_i^u \\ i \in \{0,1\} & \text{if } k < 0 \text{ and } T^k x \in Q_i^s. \end{cases}$$

It is easy to verify that ϕ is onto and that it is a.e. invertible. It remains to specify the measure on the Horseshoe, we can just pull back any invariant measure on the shift and we will get an invariant measure on the set Λ.

Let us conclude with a final remark on the physical relevance of the concept just introduced. As we mentioned, if f is an observable, then its ergodic average represents the result of an observation over a very long time (the time scale being determined by the mixing properties of the system). Yet, in reality, it may happen that we look for too short a time or, after studying a certain quantity, we can get a grant to buy the needed apparatus to perform more precise measurements. What would we see in such a case? Clearly, we would not see a constant, even for an ergodic system, and we would interpret the non constant part as fluctuations. In many cases it may happen that this fluctuations have a very special nature: they are Gaussian. In such a case we say that the system satisfies the Central Limit Theorem (CLT). Let us be more precise: define $S_n f := \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f \circ T^i$.

Definition 6.8.4 *Given a Dynamical System (X, T, μ) and a class of observables $\mathcal{A} \subset L^2(X, \mu)$ we say that the class $\mathcal{A}$ satisfies the CLT if $\forall f \in \mathcal{A}$, $\mu(f) = 0$,*

$$\lim_{n\to\infty} \mu(\{x \mid S_n f \geq t\}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2\sigma^2}} dx,$$

where (the variance) σ is defined by $\sigma^2 = \mu(f) + 2\sum_{i=1}^{\infty} \mu(f \circ T^i f)$.[39]

The relevance of the above theorem is the following: if the system is ergodic and satisfies the CLT, then $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - \mu(f) = \mathcal{O}(\frac{1}{\sqrt{n}})$, we have thus the precise scale on which the fluctuations should appear.

In this book we will be mainly interested in the question of how to establish if a given system is ergodic or not.

Unfortunately, neither ergodicity is a typical property of dynamical systems, nor is regular motion. It is a frustrating fact of life that generically dynamical systems present some kind of mixed behavior. Nevertheless, there are some class of systems that are known to be

[39]This definition is a bit stricter than usual because, in general, there may be cases in which the fluctuations are Gaussian but the formula for the variance does not hold as written.

ergodic and among them the hyperbolic systems are probably the most relevant. We will discuss them in the next chapters.

Problems

6.1. Given a measurable Dynamical Systems (X, T, μ) verify that, for each measurable set A, if $T(A)$ is measurable, then $\mu(TA) \geq \mu(A)$.

6.2. Set $\mathcal{M}^1(X) = \{\mu \in \mathcal{M} \mid \mu(X) = 1\}$ and $\mathcal{M}^1_T(X) = \mathcal{M}^1(X) \cap \mathcal{M}_T(X)$. Prove that $\mathcal{M}^1_T(X)$ and $\mathcal{M}^1(X)$ are convex sets in $\mathcal{M}(x)$.

6.3. Call $\mathcal{M}^e(X) \subset \mathcal{M}^1(X)$ the set of ergodic probability measures. Show that $\mathcal{M}^e(X)$ consists of the extremal points of $\mathcal{M}_T(X)$.

6.4. Prove that the Lebesgue measure is invariant for the rotations on $\mathbb{T}$.

6.5. Consider a rotation by $\omega \in \mathbb{Q}$, find invariant measures different from Lebesgue.

6.6. Prove that the measure μ_h defined in Examples 6.1.1 (Hamiltonian systems) is invariant for the Hamiltonian flow.

6.7. Given a Poincaré section prove that there exists $c > 0$ such that $\inf \tau_\Sigma \geq c > 0$.

6.8. Show that ν_Σ, defined in (6.2.1) is well defined.

6.9. Show that the return time τ_Σ is finite ν_Σ-a.e. .

6.10. Show that ν_Σ is T_Σ invariant. Verify that, collecting the results of the last exercises, $(\Sigma, T_\Sigma, \nu_\Sigma)$ is a Dynamical System.

6.11. something about holomorphic dynamics?

6.12. Prove that the Bernoulli measure is invariant with respect to the shift.

6.13. Let Σ_p be the set of periodic configurations of Σ. If μ is the Bernoulli measure prove that $\mu(\Sigma_p) = 0$

6.14. Consider the Bernoulli shift on $\mathbb{Z}$ and define the following equivalence relation: $\sigma \sim \sigma'$ iff there exists $n \in \mathbb{Z}$ such that $T^n\sigma = \sigma'$ (this means that two sequences are equivalent if they belong to the same orbit). Consider now the equivalence classes (the space of orbits) and choose[40] a representative from each class, call the set so obtained K. Show that K cannot be a measurable set.

6.15. Compute the transfer operator for maps of $\mathbb{T}$. Prove that $\|\mathcal{L}h\|_1 \leq \|h\|_1$.

6.16. Prove the Lasota-York inequality (6.4.5).

6.17. Prove that for each sequence $\{h_n\} \subset \mathcal{C}^{(1)}(\mathbb{T})$, with the property $\sup_{n\in\mathbb{N}} \|h_n'\|_1 + \|h_n\|_1 < \infty$, it is possible to extract a subsequence converging in L^1.

6.18. Prove Corollary 6.6.3.

6.19. Prove Theorem 6.6.5

6.20. Let $U \subset X$ of positive measure, consider

$$f_U(x) = \lim \frac{1}{n} \sum_{i=0}^{n-1} \chi_U(T^i x).$$

Show that the limit exists and that the set $A_0 := \{x \in U \mid f_U(x) = 0\}$ has zero measure.

6.21. A topological Dynamical System (X, T) is called *Topologically transitive*, if it has a dense orbit. Show that if $(\mathbb{T}^d, T, m)$ is ergodic and T is continuous, then the system is topologically transitive.

6.22. Give an example of a system with a dense orbit which it is not ergodic.

6.23. Give an example of an ergodic system with no dense orbit.

6.24. Give an example of a Dynamical Systems which does not have any invariant probability measure.

[40] Attention !!!: here we are using the *Axiom of choice*.

6.25. Show that a Dynamical Systems (X, T, μ) is ergodic if and only if there does not exists any invariant probability measure absolutely continuous with respect to μ, beside μ itself.

6.26. Prove that Birkhoff theorem implies Von Neumann theorem.

6.27. Prove that if (X, T, μ) is ergodic, then all $f \in L^1(X, \mu)$ such that $f \circ T = f$ are a.e. constant. Prove also the converse.

6.28. For each measurable set A, let

$$F_{A,n}(x) = \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x).$$

be the average number of times x visits A in the time n. Show that there exists $F_A = \lim_{n\to\infty} F_{A,n}$ a.e. and prove that, if the system is ergodic, $F_A = \mu(A)$.

6.29. Prove Proposition 6.7.2 and Proposition 6.7.3.

6.30. Show that the irrational rotations are not mixing.

6.31. Prove that if $f \in \mathcal{C}^2(\mathbb{T})$, then its Fourier series converges uniformly.[41]

6.32. Let ν be a Borel measure on $Q = [0, 1]^2$ such that $\nu(\partial_x f) = 0$ for all $f \in \mathcal{C}^1_{\text{per}}(Q) = \{f \in \mathcal{C}^1(Q) \mid f(0, y) = f(1, y)\ \forall\, y \in [0, 1]\}$. Prove that there exists a Borel measure ν_1 on $[0, 1]$ such that $\nu = m \times \nu_1$.

6.33. Prove that is a flow is ergodic (mixing) so is each Poincarè section. Prove that is a map is ergodic so is any suspension on the map. Give an example of a mixing map with a non-mixing suspension (constant ceiling).

6.34. Consider $([0, 1], T)$ where

$$T(x) = \frac{1}{x} - \left[\frac{1}{x}\right]$$

[41]This result is far from optimal, see [?] if you want to get deeper in the theory of Fourier series.

($[a]$ is the integer part of a), and

$$\mu(f) = \frac{1}{\ln 2}\int_0^1 f(x)\frac{1}{1+x}dx.$$

Prove that $([0,1], T, \mu)$ is a Dynamical System.[42]

6.35. In view of the two previous exercises explain why it is problematic to study the statistical properties of the Gauss map on a computer.

6.36. Choose a number in $[0,1]$ at random according to Lebesgue distribution. Assuming that the Gauss map is mixing (which it is, see ???) compute the average percentage of numbers larger than n in the associated continuous fraction.

6.37. Let (X_0, T_0, μ_0) be a Dynamical System and $\phi : X_0 \to X_1$ an homeomorphism. Define $T_1 := \phi \circ T_0 \circ \phi^{-1}$ and $\mu_1(f) = \mu_0(f \circ \phi^{-1})$. Prove that (X_1, T_1, μ_1) is a Dynamical System.

6.38. Let (X_0, T_0, μ_0) be measurably conjugate to (X_1, T_1, μ_1), then show that one of the two is ergodic if and only if the other is ergodic. Prove the same for mixing.

6.39. Show that the systems described in Examples ??–strange attractor and horseshoe, are Bernoulli.

6.40. Prove Lebesgue density theorem: for each measurable set A, $m(A) > 0$, there exists $x \in A$ such that for each $\varepsilon > 0$ exists $\delta > 0$ such that $m(A \cap [x-\delta, x+\delta]) > (1-\varepsilon)2\delta$.

Hints to solving the Problems

6.3 Use Krein-Milman Theorem [DS88].

6.6 Use the properties of H to deduce $\langle \nabla_{\phi^t x} H,\, d_x\phi^t \nabla_x H\rangle = \|\nabla_x H\|^2$, and thus $d_x\phi^t\nabla_x H = \frac{\|\nabla_x H\|^2}{\|\nabla_{\phi^t x} H\|^2}\nabla_{\phi^t x}H + v$ where $\langle \nabla_{\phi^t x} H,\, v\rangle = 0$.

[42] The above map is often called *Gauss map* since to him is due the discovery of the above invariant measure.

Then study the evolution of an arbitrarily small parallelepiped with one side parallel to $\nabla_x H$–or look at the volume form if you are more mathematically incline–remembering the invariance of the volume with respect to the flow.

6.8 Use the invariance of μ and the fact that, by Problem 6.7, if $A \subset \Sigma$ then $\mu(\phi^{[0,\delta]}(A) \cap \phi^{[n\delta,\,(n+1)\delta]}A) = 0$ provided $(n+1)\delta \leq c$.

6.9 Let $\delta < c$ and $\Sigma_\delta := \phi^{[0,\delta]}\Sigma$, apply Poincaré return theorem to Σ_δ.

6.12 Check it on the algebra $\mathcal{A}$ first.

6.13 Σ_p is the countable union of zero measure sets.

6.14 Show that $K \cap T^n K \subset \Sigma_p$, then by using Problem 6.13 show that if K is measurable $\sum_{i=-\infty}^{\infty} \mu(T^n K) = 1$ which, by the invariance of μ, is impossible.

6.15 Use the equivalent definition $\int g\mathcal{L}f dm = \int fg \circ T dm$.

6.17 Consider partitions $\mathcal{P}_n$ of $\mathbb{T}$ in intervals of size $\frac{1}{n}$. Define the conditional expectation $\mathbb{E}(h|\mathcal{P}_n)(x) = \frac{1}{m(I(x))}\int_{I(x)} hdm$, where $x \in I(x) \in \mathcal{P}_n$. Prove that $\|\mathbb{E}(h|\mathcal{P}_n) - h\|_1 \leq \frac{1}{n}\|h'\|_1$. Notice that the functions $\mathbb{E}(h_n|\mathcal{P}_m)$ have only m distinct values and, by using the standard diagonal trick, construct an subsequence h_{n_j} such that all the $\mathbb{E}(h_{n_j}|\mathcal{P}_m)$ are converging. Prove that h_{n_j} converges in L^1.

6.19 Note that $\mu(T^{-n}A \cap T^{-m}A) \neq 0$ then, supposing without loss of generality $n < m$, $\mu(A \cap T^{-m+n}A) \neq 0$. Then prove the theorem by absurd remembering that $\mu(X) < \infty$.

6.20 The existence follows from Birkhoff theorem, it also follows that A_0 is an invariant set, then

$$0 = \int_{A_0} f_U = \int_{A_0} \chi_U = \mu(A_0).$$

6.21 For each $n \in \mathbb{N}$, $x \in \mathbb{T}^d$ consider $B_{\frac{1}{m}}(x)$–the ball of radius $\frac{1}{m}$ centered at x. By compactness, there are $\{x_i\}$ such that $\cup_i B_{\frac{1}{m}}(x_i) = \mathbb{T}^d$. Let

$$A_{m,i} = \{y \in \mathbb{T}^d \mid T^k y \cap B_{\frac{1}{M}}(X_I) = \emptyset \ \ \forall k \in \mathbb{N}\},$$

clearly $A_{m,i} = \cap_{k\in\mathbb{N}} T^{-k} B_{\frac{1}{m}}(x_i)^c$ has the property $T^{-1}A_{m,i} \supset A_{m,i}$. It follows that $\tilde{A}_{m,i} = \cup_{n\in\mathbb{N}} T^{-n} A_{m,i} \supset A_{m,i}$ is an invariant set and it holds $\mu(\tilde{A}_{m,i} \backslash A_{m,i}) = 0$. Since $A_{m,i}$ it is not of full measure, $\tilde{A}_{m,i}$, and thus $A_{m,i}$, must have zero measure. Hence, $\bar{A}_m = \cap_i A_{m,i}$ has zero measure. This means that $\cup_{m\in\mathbb{N}} \bar{A}_m$ has zero measure. Prove now that, for each $y \in \mathbb{T}^d$, the trajectories that never get closer than $\frac{2}{m}$ to y are contained in $\bar{A}_m$, and thus have measure zero. Hence, almost every point has a dense orbit.) Extend the result to the case in which X is a compact metric space and μ charges the open sets (that is: if $U \subset X$ is open, then $\mu(U) > 0$.

6.22 A system with two periodic orbits, and the measure supported on them. Along such lines more complex examples can be readily constructed.

6.23 A non transitive system with a measure supported on a periodic orbit.

6.24 $X = \mathbb{R}^d$, $Tx = x + v$, $v \neq 0$.

6.26 Note that the ergodic average is a contraction in L^∞, an isometry in L^2 and that $L^1 \subset L^2$ (since the measure is finite). Use Lebesgue dominate convergence theorem to prove convergence in L^2 for bounded functions. Use Fatou to show that if $f \in L^2$ then $f^+ \in L^2$ and a 3-ε argument to conclude.

6.28 Birkhoff theorem and Theorem 6.6.6.

6.29 Note that for each measurable set A and $\varepsilon > 0$ there exists $f \in \mathcal{C}^0(X)$ such that $\mu(|f - \chi_A|) < \varepsilon$ –by Uryshon Lemma and by the regularity of Borel measures. To prove that $\mu(T^{-n}A \cap$

$B) \to \mu(A)\mu(B)$ choose $d\lambda = \mu(B)^{-1}\chi_B d\mu$ and use the invariance of μ to obtain the uniform estimate $\lambda(|f \circ T^n - \chi_A \circ T^n|) \leq \mu(B)^{-1}\mu(|f - \chi_A|)$.

6.31 Remember that $f_n = \frac{1}{2\pi}\int_{\mathbb{T}} e^{2\pi i n x} f(x)dx$. Thus

$$f_n = \frac{1}{(2\pi i n)^2 2\pi}\int_{\mathbb{T}} e^{2\pi i n x} f^{(2)}(x)dx.$$

6.32 The measure ν_1 is nothing else then the marginal with respect to x, that is: for each continuous function $f : [0,1] \to \mathbb{R}$ define $\tilde{f} : Q \to \mathbb{R}$ by $\tilde{f}(x,y) = f(y)$, then $\nu_1(f) = \nu(\tilde{f})$. To prove the statement use Fourier series. If f is smooth enough $f(x,y) = \sum_{k\in\mathbb{Z}} \hat{f}_k(y)e^{2\pi i k x}$ where the Fourier series for f and $\partial_x f$ converge uniformly. Then notice that $0 = \nu(\partial_x e^{2\pi i k \cdot}) = 2\pi i k \nu(e^{2\pi i k \cdot})$ implies $\nu(f) = \nu(\hat{f}_0) = m \times \nu_1(f)$.

6.34 Write $\mu(f \circ T) = \sum_{i=1}^{\infty}\int_{\frac{1}{i+1}}^{\frac{1}{i}} f \circ T(x)\mu(dx)$, change variable and use the identity $\frac{1}{a^2+a} = \frac{1}{a} - \frac{1}{a+1}$ to obtain a series with alternating signs.

6.35 The computer uses only rational numbers. It is quite amazing that these type of pathologies arises rather rarely in the numerical studies carried out by so many theoretical physicist.

6.36 Define $f(x) = [x^{-1}]$, then the entries of the continuous fraction of x are $\{f \circ T^i\}$. The quantity one must compute is then $m(\lim_{k\to\infty} \frac{i}{k}\sum_{i=0}^{k-1} \chi_{[n,\infty)} \circ f \circ T^i) = \mu([n,\infty))$.

6.40 We have seen in Examples 6.8.1-Dilations that Lebesgue measure is equivalent to Bernoulli measure and that the cylinder correspond to intervals. It then suffices to prove the theorem for the latter. Let $A \subset \Sigma^+$ such that $\mu(A) > 0$, then, for each $\varepsilon > 0$,there exists $A_\varepsilon \in \mathcal{A}$ such that $A_\varepsilon \supset A$ and $\mu(A_\varepsilon) - \mu(A) < \varepsilon\mu(A)$. Since $A_\varepsilon \in \mathcal{A}$, it exists $n_\varepsilon \in \mathbb{N}$ such that it is possible to decide if $\sigma \in A_\varepsilon$ only by looking at $\{\sigma_1, \dots, \sigma_{n_\varepsilon}\}$. Consider all the cylinders $\mathcal{I}\{A(0; k_1, \dots, k_{n_\varepsilon})\}$, clearly if $I \in \mathcal{I}$ then $I \cap A_\varepsilon$ is either I or $\emptyset$. Let $\mathcal{I}_+ = \{I \in \mathcal{I} \mid I \cap A_\varepsilon = I\}$ and $\mathcal{I}_+ = \{I \in \mathcal{I} \mid I \cap A_\varepsilon = \emptyset\}$.

Now suppose that for each $I \in \mathcal{I}_+$ holds $\mu(I \cap A) \leq (1-\varepsilon)\mu(I)$ then

$$\mu(A) = \sum_{I \in \mathcal{I}_+} \mu(A \cap I) \leq (1-\varepsilon)\mu(A_\varepsilon) < \mu(A),$$

which is absurd. Thus there must exists $I \in \mathcal{I}_+$: $\mu(A \cap I) > (1-\varepsilon)\mu(I)$.

Notes

Give references for SRB and Gibbs, mention entropy, K-systems. diffeo with holes, strange attractors, history of the field

. . . .