BERNOULLI SCHEMES AND THEIR ISOMORPHISMS

Giovanni Gallavotti

Lecture I

In these lectures I shall present some of the ideas and methods used in the proof of the celebrated theorem stating that two Bernoulli schemes are isomorphic if and only if they have the same entropy.

I shall concentrate, because of lack of time, on a well known partial result, Sinai's theorem, which states that two Bernoulli schemes are weakly isomorphic if and only if they have the same entropy.

In these lectures I shall present some of the ideas and methods used in the proof of the celebrated theorem stating that two Bernoulli schemes are isomorphic if and only if they have the same entropy.

I shall concentrate, because of lack of time, on a well known partial result, Sinai's theorem, which states that two Bernoulli schemes are weakly isomorphic if and only if they have the same entropy.

I shall outline Ornstein's proof of Sinai's theorem which gives an insight into the coding problems that arise in connection with the isomorphism problems.

Let me start with some basic notions.

A Bernoulli scheme is a mathematical model for a "head and tail" game: we have k symbols $(1, 2, \ldots, k)$ and a machine that extracts out of a box one of these symbols, randomly, with respective probability $p_1, p_2, \ldots, p_k$ and so that successive extractions are independent.

The output of the machine is a sequence $\{x_i\}_{-\infty}^{+\infty}$, where $x_i = 1, 2, \ldots, k$.

We ask the question of which is the probability that the machine produces a particular sequence.

To make precise this question we proceed as usual, in probability theory:

1. Definition: *The space $\mathcal{K}$ consisting in all sequences $\{x_i\}_{-\infty}^{+\infty}, x_i = 1, 2, \ldots, k$, is called configuration space. The subsets $E(\,{}^{i_1,i_2,\ldots,i_\ell}_{x_1,x_2,\ldots,x_l}\,)$ consisting in the sequences whose entries with index ("time") $i_1, i_2, \ldots, i_l$ are $x_1, x_2, \ldots, x_l$ will be called "cylinders" in $\mathcal{K}$.*

It is clear that the probability that the machine output is in $E(\,{}^{i_1,i_2,\ldots,i_\ell}_{x_1,x_2,\ldots,x_l}\,)$, $i_1 < \ldots < i_p$, should be

e1.1

$$\mu\Big(E\big(\begin{matrix} i_1, i_2, \ldots, i_\ell \\ x_1, x_2, \ldots, x_l \end{matrix}\big)\Big) \stackrel{def}{=} \prod_{i=1}^{\ell} p_{x_i} \tag{1.1}$$

Once we have defined the probability of the cylinders it is possible to extend it to a probability measure μ defined on the smallest σ-algebra of subsets of $\mathcal{K}$ which contains the cylinders. This is very easy to check in our special case but it is a particular result of a more general theorem (also easy to prove in its simple version given below) which we state now for future reference.

Assume that with every cylinder $E(\,{}^{i_1,i_2,\ldots,i_\ell}_{x_1,x_2,\ldots,x_l}\,)$, $i_1 < \ldots < i_p$, in $\mathcal{K}$ is associated a number $\mu\Big(E(\,{}^{i_1,i_2,\ldots,i_\ell}_{x_1,x_2,\ldots,x_l}\,)\Big)$ and

e1.2

$$\begin{aligned}
&\mu\Big(E\Big(\begin{matrix} i_1,i_2,\ldots,i_\ell \\ x_1,x_2,\ldots,x_l \end{matrix}\Big)\Big) \ge 0 \quad \forall i_1,\ldots,i_\ell, \quad \forall x_1,\ldots,x_\ell;\ell \\
&\sum_{x_1,\ldots,x_\ell}^{1,k} \mu\Big(E\Big(\begin{matrix} i_1,i_2,\ldots,i_\ell \\ x_1,x_2,\ldots,x_l \end{matrix}\Big)\Big) = 1 \;\forall \ell,\; \forall i_1,\ldots,i_\ell \\
&\sum_{y_1,\ldots,y_p} \mu(E\Big(\begin{matrix} i_1,\ldots,i_\ell,j_1,\ldots,j_p \\ x_1,\ldots,x_\ell,y_1,\ldots,y_p \end{matrix}\Big)) = \mu\Big(E\Big(\begin{matrix} i_1,i_2,\ldots,i_\ell \\ x_1,x_2,\ldots,x_l \end{matrix}\Big)\Big)
\end{aligned} \tag{1.2}$$

$\forall i_1,\ldots,i_\ell; x_1,\ldots,x_\ell; j_1,\ldots,j_p$ such that and $j_s \ne i_r$ $(s = 1,\ldots,p; r = 1,\ldots,\ell)$.

Then the following theorem (Kolmogorov) holds

I. Theorem: *there is a unique measure μ defined on the smallest σ-algebra containing the cylinders which assigns $\mu\Big(E(\,{}^{i_1,i_2,\ldots,i_\ell}_{x_1,x_2,\ldots,x_l}\,)\Big)$ as measure (that we call probability) of $E(\,{}^{i_1,i_2,\ldots,i_\ell}_{x_1,x_2,\ldots,x_l}\,)$, $\forall \ell, i_1,\ldots,i_\ell; x_1,\ldots, x_\ell$.*

Of particular interest will be the measures, defined on the smallest σ- algebra containing the cylinders and which have the "shift" invariance property.

e1.3

$$\mu\Big(E\Big(\begin{matrix} i_1+1, i_2+1, \ldots, i_\ell+1 \\ x_1 \;\;,\;\; x_2 \;\;,\ldots,\;\; x_l \end{matrix}\Big)\Big) \equiv \mu\Big(E\Big(\begin{matrix} i_1,i_2,\ldots,i_\ell \\ x_1,x_2,\ldots,x_l \end{matrix}\Big)\Big) \tag{1.3}$$

$\forall i_1,\ldots,i_\ell, x_1,\ldots,x_\ell$. If μ is a measure on the smallest σ-algebra Σ containing the cylinders, we shall always imagine it extended to the σ-algebra Σ_μ obtained by adding to Σ the sets with outer measure zero, also called "null sets": in general a measure μ defined on a σ-algebra containing all sets with outer measure zero is called complete, and the σ-algebra is called μ-complete.

The reason why we consider the latter extension is the following theorem

II. Theorem: *A complete non atomic measure μ defined on a μ-countably generated σ-algebra is isomorphic to the Lebesgue measure on the unit interval.*

"Non atomic" means that every measurable set with > 0 measure contains a measurable subset with smaller but > 0 measure.

The σ-algebra Σ_μ over which μ is defined is said to be "μ-countably generated" if it is obtained by adding the sets of outer μ-measure zero to a σ-algebra which is the smallest σ-algebra containing a denumerable family of sets.

"Isomorphic" means that there is a mapping ψ which maps equivalence classes of μ-measurable sets (with the equivalence relation $E \sim F$ if $\mu(E\Delta F) = 0$) of $\mathcal{K}$ onto equivalence classes of Lebesgue measurable sets in $[0,1]$ (where the equivalence is now with respect to the Lebesgue measure) in such a way that $\psi(\mathcal{K} - E) = \psi(\mathcal{K}) - \psi(E), \psi\big(\cup_{i=1}^\infty \psi(E_i)\big) = \cup_{i=1}^\infty \psi(E_i)$, for all measurable sets $E, E_1, E_2, \ldots$, and $\mu(E) =$ (Lebesgue measure of $\psi(E)$).

The interest of the above theorem lies in its implication that if $E \subset \mathcal{K}$ is μ-

measurable and $0 < \alpha < \mu(E), \exists A \subset E$ such that $\mu(A) = \alpha$.

2. Definition: *The triple $(\mathcal{K}, \mu, T)$ formed by a non atomic probability distribution of the type described in theorem 2 defined on a space $\mathcal{K}$ of sequences and invariant under the shift T of the sequences will be called a "shift". The shift T has to be thought as the map $\{x_i\}_{-\infty}^{+\infty} \to \{x'_i\}_{-\infty}^{+\infty}$ with $x'_i = x_{i+1}$. More generally a "dynamical system" is a triple $(\mathcal{K}, \mu, T)$ consisting in a measure $(\mathcal{K}, \mu)$ isomorphic to the Lebesgue measure on [0,1] and T is an invertible transformation of $\mathcal{K} - \mathbb{N}$ onto $\mathcal{K} - \mathbb{N}$ mapping in a measure preserving way the measurable subsets of $\mathcal{K} - \mathbb{N}$ onto the measurable subsets of $\mathcal{K} - \mathbb{N}$ where $\mathbb{N}$ is a suitable null set.*

Given an abstract dynamical system it is possible to construct many shifts: let $\mathcal{P} = (P_{1,\ldots,}P_n)$ a partition of $\mathcal{K}$ (with measurable sets).

Let $\Psi : \mathcal{K} \to \widetilde{\mathcal{K}}$ be the mapping of $\mathcal{K} - \mathbb{N}$ into the space $\widetilde{\mathcal{K}}$ of sequences of k symbols $(1, \ldots, k)$ defined by $\Psi x = \{x_i\}_{-\infty}^{+\infty}$ where x_i is such that $T^i x \in P_{x_i}$. Clearly the transformation T is mapped, in a natural sense, into the shift $\widetilde{T}$ on $\widetilde{\mathcal{K}}$. Furthermore if we put

e1.4

$$\widetilde{\mu}(E^{i_1 \ldots i_\ell}_{x_1 \ldots x_\ell}) = \mu(\bigcap_{s=1}^{\ell} T^{-s} P_{x_s}) \tag{1.4}$$

we can check that $\widetilde{\mu}$ verifies the conditions of Kolmogorov's theorem and, therefore, defines a measure $\widetilde{\mu}$ on $\widetilde{\mathcal{K}}$.

There is also another natural dynamical system that can be associated with a partition $\mathcal{P}$ of $\mathcal{K}$.

We need some definitions and conventions: every partition will always be regarded as an ordered partition (i.e. if $\mathcal{P} = (P_1, P_2)$ then $\mathcal{P}' = (P_2, P_1) \neq \mathcal{P}$); if $\mathcal{P}, \mathcal{Q}$ are partitions of the same space we can consider the partition whose elements ("atoms") are of the form $P_i \bigcap Q_j$ with $P_i \in \mathcal{P}, Q_j \in \mathcal{Q}$ ordered lexicographically: this partition will be denoted $\mathcal{P} \vee \mathcal{Q}$. If $(\mathcal{K}, \mu, T)$ is a dynamical system and $\mathcal{P}$ is a partition of $\mathcal{K}$ we can define $\vee_{i=-\infty}^{\infty} T^i \mathcal{P}$ to be the smallest complete σ-algebra which contains the atoms of $\vee_{i=-N}^{N} T^i \mathcal{P}, \forall N \geq 0$.

3. Definition: *If $(\mathcal{K}, \mu, T)$ is a dynamical system and $\mathcal{P}$ is a measurable partition of $\mathcal{K}$ we shall denote $(T, \mathcal{P})$ the dynamical system $(\mathcal{K}, \overline{\mu}, T)$ where $\overline{\mu}$ is obtained by restricting μ to the σ-algebra $\vee_{i=-\infty}^{\infty} T^i \mathcal{P}$.*

It is legitimate to suspect that the dynamical system $(\mathcal{P}, T)$ and the shift $(\widetilde{\mathcal{K}}, \widetilde{\mu}, \widetilde{T})$ associated with $\mathcal{P}$ in the previous construction are, in fact, the same thing. Therefore the following definition of isomorphism is very natural.

4. Definition: *Two dynamical systems $(\mathcal{K}, \mu, T)$ and $(\mathcal{K}', \mu', T')$ are isomorphic if $(\mathcal{K}, \mu)$ and $(\mathcal{K}', \mu')$ are isomorphic as measure spaces (c.f.r. theorem 3) and if the isomorphism Ψ of $(\mathcal{K}, \mu)$ onto $(\mathcal{K}', \mu')$ commutes with T, T': i.e. $T'\Psi = \Psi T$ (where ΨT and $T'\Psi$ are regarded as transformations of equivalence classes of measurable sets).*

One then checks that $(T, \mathcal{P})$ is indeed isomorphic to the shift $(\widetilde{\mathcal{K}}, \widetilde{\mu}, \widetilde{T})$ constructed from $(\mathcal{K}, \mu, T)$ using the partition $\mathcal{P}$ as described above.

Since the shift $(\widetilde{\mathcal{K}},\widetilde{\mu},\widetilde{T})$ is completely determined by the knowledge of the numbers $\widetilde{\mu}(E^{i_1,\ldots,i_\ell}_{x_1,\ldots,x_\ell})$ the process $(T,\mathcal{P})$ will be completely determined, up to isomorphism, by the joint distributions $\mu(\cap_{s=1}^{\ell}T^{-i_s}P_{x_{i_s}})$, i.e. by the probabilities of the atoms of the partitions $\vee_{i=0}^{N-1}T^{-i}\mathcal{P}$.

We shall say that $(T,\mathcal{P})$ and $(T',\mathcal{P}')$ are "copies" if $\mathcal{P},\mathcal{P}'$ contain the same number of elements and

e1.5

$$\mu(\bigcap_{s=1}^{\ell} T^{-i_s}P_{x_s}) = \mu'(\bigcap_{s=1}^{\ell} T'^{-i_s}P'x_s) \tag{1.5}$$

for all $\ell, i_1,\ldots,i_\ell, x_1,\ldots,x_\ell$. Clearly two copies are isomorphic as dynamical system. Some dynamical systems admit what is called a "generating" partition $\mathcal{P} = (P_1,\ldots P_k)$: i.e., denoting $\overline{\Sigma}$ the completion of a σ-algebra Σ with respect to μ, a partition such that $\Sigma_\mu = \overline{\vee_{i=-\infty}^{\infty}T^i\mathcal{P}}$. It can be shown that, when Σ_μ is the completion of the Borel σ-algebra of a separable metric space $\mathcal{K}$, such partitions are characterized, when they exist, by the property that μ- almost all $x\in\mathcal{K}$ are determined uniquely by the sequence $\{x_i\}_{-\infty}^{+\infty}$. In this case $(\mathcal{K},\mu,T)$ and $(T,\mathcal{P})$ are isomorphic. If $(\mathcal{K},\mu,T)$ is a shift there is a natural generating partition: namely the partition $\mathcal{P}$ which partitions $\mathcal{K}$ according to the value of the entry in $\{x_i\}_{-\infty}^{+\infty}\in\mathcal{K}$ with index 0. The systems $(\mathcal{K},\mu,T)$ and $(\mathcal{P},T)$ are in this case isomorphic. The general problem solved by Ornstein is the one of characterising those dynamical systems which are isomorphic to a Bernoulli scheme or, what amounts to the same, those dynamical systems which have a generating partition which is "independent", i.e. such that

e1.6

$$\mu(\cap_{\ell=1}^{s}T^{\ell}P_{x_\ell}) = \prod_{\ell=1}^{s}\mu(P_{x_\ell}) \tag{1.6}$$

We shall be interested, in these lectures, in an easier problem: namely the one studying when two Bernoulli schemes are weakly isomorphic: the notion of weak isomorphism is the following:

5. Definition: *Two shifts $(\mathcal{K},\mu,T)$ and $(\mathcal{K}',\mu',T')$ are weakly isomorphic if there are partitions $\mathcal{P}$ of $\mathcal{K}$ and $\mathcal{P}'$ of $\mathcal{K}'$ such that $(\mathcal{K},\mu,T)$ is isomorphic to $(\mathcal{P}',T')$ and $(\mathcal{K}',\mu',T')$ is isomorphic to $(\mathcal{P},T)$.*

In other words two dynamical systems are weakly isomorphic if each of them is a "subsystem of the other".

Finally we need the definition of ergodicity since we shall assume that all the dynamical systems we consider in these lectures are ergodic:

6. Definition: *A dynamical system is ergodic if there are no invariant sets with measure different from zero or one.*

Exercise: shows that a Bernoulli scheme is ergodic and non atomic.

Lecture II

Let $(\mathcal{K}, \mu, T)$ be a shift.

We start our discussion by remarking that not all the possible sequences will have an appreciable probability of appearing.

This intuitive notion can be made precise in several ways.

Let us consider first the simple case when $(\mathcal{K}, \mu, T)$ is a Bernoulli scheme with symbols $(1, 2, \ldots k)$ and probabilities $(p_1, \ldots, p_k)$.

Consider the cylinders of length N (i.e. look at the first N outputs of the machine which generates the "game").

If $E^{1,\ldots,N}_{x_1,\ldots,x_N}$ is one such cylinder we can write its probability:

e2.1 $$\mu(E^{1,\ldots,N}_{x_1,\ldots,x_N}) = \prod_{i=1}^{N} p_{x_i} = \exp \sum_{1=i}^{N} \log p_{x_i} \tag{2.1}$$

Remark that if $f : \mathcal{K} \to R$ is the function on $\mathcal{K}$ defined as

e2.2 $$f(\{x_i\}^{+\infty}_{-\infty}) = \log p_{x_0} \tag{2.2}$$

the sum in the above exponential can be expressed as N times

e2.3 $$\frac{1}{N} \sum_{i=1}^{N} f(T^{-i}x) \qquad \forall\, x \,\in\, E({}^{1,\ldots,N}_{x_1,\ldots,x_N}) \tag{2.3}$$

(where we use the same symbol for the shift operating on the sets and on the points).

To estimate $\mu(E^{1,\ldots,N}_{x_1,\ldots,x_N})$ we can make use of the Birkoff-VonNeumann theorem

III. Theorem: *let $(\mathcal{K}, T, \mu)$ be an ergodic shift and let $f \in L_\infty(\mathcal{K}, \mu)$ then*

e2.4 $$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(T^i x) = \int_{\mathcal{K}} f(x')\mu(dx') \tag{2.4}$$

for almost all $x \in \mathcal{K}$. This limit is therefore also reached in the L_1 sense and in μ-measure.

We recall that convergence in measure can be expressed by saying that given $\varepsilon > 0$ there exists N_ε such that $\forall\, N \,\geq N_\varepsilon$ the set

e2.5 $$\{x \,|\, |N^{-1} \sum_{i=1}^{N} f(T^i x) - \overline{f}| \,>\, \varepsilon\} \tag{2.5}$$

has measure $< \;\varepsilon$. If we apply the above theorem to our Bernoulli scheme using

e2.6 $$f(\{x_i\}^{+\infty}_{\infty}) = \log p_{x_0} \quad \text{and} \quad \int_{\mathcal{K}} f\, d\mu = \sum_{i=1}^{k} p_i \log p_i \;=\; -s \tag{2.6}$$

we see that the set of the $x's$ such that

e2.7
$$|N^{-1}\sum_{i=1}^{N}\log p_{x_i}+s|>\varepsilon \tag{2.7}$$

has measure $<\varepsilon$ for $N\ge N_\varepsilon$. It is clear that such a set is a union of cylinders of the form $E^{1,\ldots,N}_{x_1,\ldots x_N}$ (since $N^{-1}\sum_{i=1}^{N}f(T^ix)$ only depends on $x_1,\ldots,x_N$ if $x=\{x_i\}^{+\infty}_{\infty}$).

Therefore we can divide the cylinders with base $(1,2,\ldots,N)$ in two classes:
(1) the first class consists of a family of cylinders whose union has measure $\le\varepsilon$
(2) to the second class belong cylinders $E^{1,\ldots,N}_{x_1,\ldots,x_N}$ such that

e2.8
$$|N^{-1}\sum_{i=1}^{N}\log p_{x_i}+s|<\varepsilon \tag{2.8}$$

i.e. "all but ε cylinders" have measure between

e2.9
$$\exp\ N\,(-s\pm\varepsilon) \tag{2.9}$$

if $N\ge N_\varepsilon$. The number of cylinders in this class clearly cannot exceed $\exp N(s+\varepsilon)$.

This is the type of result we were looking for: if we consider a string of N events and N is large enough the number of strings that we can "reasonably" expect to see is $\sim\exp s\,N$ out of $k^N=\exp N\,\log k$ that are a priori possible (note that $s\le\log k$): furthermore all the "possible" strings have roughly the same probability $\exp -s\,N$.

The above discussion can be generalized to a general ergodic shift on a space of sequences (Shannon-McMillan):

IV. Theorem: *Let $(\mathcal{K},\mu,T)$ be an ergodic shift over a space of sequences of k symbols $(1,2,\ldots,k)$. Given $\varepsilon>0$, $\exists N_\varepsilon$ such that if $N>N_\varepsilon$ the cylinders can be divided into two classes $\mathcal{C}^1_N,\mathcal{C}^2_N$ such that*

e2.10
$$\begin{aligned}&(1)\ \mu\big(\bigcup_{E\in\mathcal{C}^1_N}E\big)\ <\ \varepsilon\\ &(2)\ \text{if } E\ \in\ \mathcal{C}^2_N\quad\text{then } \mu(E)\text{ is between } e^{-(s\pm\varepsilon)N}\end{aligned} \tag{2.10}$$

where s is a suitable non negative number called "entropy" of the shift.

The proof of this theorem is analogous to the one given for the Bernoulli case and it is very instructive when $(\mathcal{K},\mu,T)$ is a "Markov process", i.e. when $\forall N>0,\forall x_0,x_{\pm1},\ldots,x_{\pm N}$

e2.11
$$\frac{\mu(E^{-1\,0\,1}_{x_{-1}x_0x_1})}{\mu(E^{-1\,1}_{x_{-1}x_1})}=\frac{\mu(E^{-N\,\ldots\,-1\,0\,1\,\ldots\,N}_{x_{-N}\,\ldots\,x_{-1}\,x_0\,x_1\,\ldots\,x_N})}{\mu(E^{-N\ldots\,-1\,1\,\ldots\,N}_{x_{-N}\ldots x_{-1}\,x_1\ldots x_N})} \tag{2.11}$$

It might seem that if we construct a new ergodic shift $(\mathcal{P},T)$ by choosing a generating partition of the ergodic shift the entropy of the new shift would be different from the entropy of the original shift.

This is not the case and it is the content of the following theorem (Sinai):

V. Theorem: *all shifts $(\mathcal{P},T)$ obtained from an ergodic shift by choosing a generating $\mathcal{P}$ have the same entropy.*

This implies that two ergodic shifts with different entropy cannot be isomorphic.

To obtain a better insight into the isomorphism problem let us strengthen the above Shannon–McMillan theorem by combining it with the Birkoff–VonNeuman theorem.

Let $(\mathcal{K},\mu,T)$ be an ergodic shift on sequences $\{x_1\}_{-\infty}^{+\infty}$, $x_i \in (1,2,\ldots,k) = I$. let $u > 0$ be a fixed integer and let $(x_1,\ldots,x_u)$ be a sequence of symbols in I. Let $E^{1,\,\ldots,\,N}_{y_1,\ldots,y_N}$ be a cylinder in $\mathcal{K}$; we can define the *frequency* $\alpha(x_1,\ldots,x_u)$ of $(x_1,\ldots,x_u)$ in $E^{1,\,\ldots,\,N}_{y_1,\ldots,y_N}$ as: the number of $\tau's$, with $1 \le \tau \le N-u$, such that $y_\tau = x_1, y_{\tau+1} = x_2, \ldots y_{\tau+u-1} = x_u$ divided by N.

If we define χ to be the characteristic function of the cylinder $E^{1,\,\ldots,\,u}_{x_1,\ldots,x_u}$ we see that the above frequency is

e2.12
$$N^{-1}\sum_{i=0}^{N-u}\chi(T^{-i}x) \qquad x \in E^{1,\,\ldots,\,N}_{y_1,\ldots,y_N} \tag{2.12}$$

The Birkoff–VonNeuman theorem tells us that this frequency tends, in measure, to

e2.13
$$\int \chi\, d\,\mu = \mu(E^{1,\,\ldots,\,u}_{x_1,\ldots,x_u}) \tag{2.13}$$

This remark allows us to formulate the following, only apparently stronger, version of the Shannon–McMillan theorem:

VI. Theorem: *Let $(\mathcal{K},\mu,T)$ be an ergodic shift over a space of sequences of symbols in $I=(1,2,\ldots,k)$. There exists $s \ge 0$, called the "entropy", such that given $\varepsilon > 0, u > 0$ there exists $N_{\varepsilon,u}$ with the property that if $N > N_{\varepsilon,u}$ the cylinders $E^{1,\,\ldots,\,N}_{x_1,\ldots,x_N}$ can be divided in two classes $\mathcal{C}^1_N, \mathcal{C}^2_N$ with*
(1) $\mu(\cup_{E\in\mathcal{C}^1_N} E) < \varepsilon$.
(2) If $E \in \mathcal{C}^2_N$ then $\mu(E)$ is between $\exp -(s \pm \varepsilon)N$.
(3) The frequency $\nu_E(x_1,\ldots,x_u)$ of the strings $(x_1,\ldots,x_u)$ in $E \in \mathcal{C}^2_N$ is such that

e2.14
$$\sum_{x_1,\ldots,x_u} |\nu_E(x_1,\ldots,x_u) - \mu(E^{1,\,\ldots,\,u}_{x_1,\ldots,x_u})| < \varepsilon \tag{2.14}$$

Lecture III

In this lecture some of the basic properties of the entropy will be recalled. We shall also state some other auxiliary results and definitions.

From the proof of the Shannon-McMillan theorem it emerges that the number s, the entropy, is given by

e3.1
$$s = \lim_{N\to\infty} -\frac{1}{N}\sum_{y_1\ldots y_N} \mu(E^{1,\ldots N}_{y_1\ldots y_N})\ \log\ \mu\ (E^{1,\ldots,N}_{y,\ldots,y_N}) \tag{3.1}$$

such a formula can also be immediately derived from the text of the theorem itself.

To list the properties of $\mathcal{S}$ it is useful to introduce the following definitions concerning measurable partitions $\mathcal{P}, \mathcal{Q}$, of a measure space (K,μ).

7. Definition: *If $\mathcal{P}, \mathcal{Q}$ are measurable partitions of a measure space (K,μ) we set* e3.2 $$\mathcal{S}(\mathcal{P})\stackrel{def}{=}-\sum_{P\in\mathcal{P}}\mu\,(P)\,\log\mu\,(P),\qquad \mathcal{S}(\mathcal{P}/\mathcal{Q})\stackrel{def}{=}\mathcal{S}(\mathcal{P}\bigvee\mathcal{Q})-\mathcal{S}(\mathcal{Q}) \tag{3.2}$$

note that $\mathcal{S}(\mathcal{P})\geq 0$.

The following theorem collects all the results we need about the entropy:

VII. Theorem: *Let* $(\mathcal{K},\mu,T)$ *be a dynamical system and* $\mathcal{P},\mathcal{Q},\mathcal{R}$ *be measurable partitions of* $\mathcal{K}$*. Denote* $\mathcal{S}(\mathcal{P},T)$ *the entropy of the process* $(\mathcal{P},T)$ *(regarded as a shift):*

(1) $\mathcal{S}\big((\mathcal{P}\bigvee\mathcal{Q})/\mathcal{R}\big)\leq\mathcal{S}(\mathcal{P}/\mathcal{R})+\mathcal{S}(\mathcal{Q}/\mathcal{R}),\qquad \mathcal{S}((\mathcal{P}\bigvee\mathcal{Q})/\mathcal{R})\geq\mathcal{S}(\mathcal{P}/\mathcal{R})$

(2) $\mathcal{S}\big(\mathcal{P}/(\mathcal{Q}\bigvee\mathcal{R})\big)\leq\mathcal{S}\,(\mathcal{P}/\mathcal{Q})$

(3) $\mathcal{S}(\mathcal{P},T)=\lim_{N\to\infty}\mathcal{S}(\mathcal{P}/\bigvee_{i=0}^{N}T^{-i}\mathcal{P})$

e3.3 (4) $\mathcal{S}(\mathcal{P},T)=\lim_{N\to\infty}N^{-1}\mathcal{S}(\bigvee_{i=0}^{N-1}T^{-i}\mathcal{P})$ (3.3)

(5) $\mathcal{S}(\bigvee_{a}^{b}T^{i}\,\mathcal{P},T)\equiv\mathcal{S}(\mathcal{P},T)$

(6) If $\mathcal{P}=(P_1,\ldots,P_k)\,,\,\mathcal{Q}=(Q_1,\ldots,Q_k)$ and $\varepsilon\stackrel{def}{=}\Delta(\mathcal{P},\mathcal{Q})=\sum_{i=1}^{k}\mu(P_i\Delta Q_i)$

then $|\mathcal{S}(\mathcal{P},T)-\mathcal{S}(\mathcal{Q},T)|\leq-(\varepsilon\log\varepsilon+(1-\varepsilon)\log(1-\varepsilon))\,+\,\varepsilon\,\log k$

Point (6) of this theorem is the key to the proof of the theorem 3 of lecture II.

Let us introduce some more elementary notions about partitions:

8. Definition: *Let* $\mathcal{P}=(P_1,\ldots,P_k),\mathcal{Q}=(R_1,\ldots,R_k)$ *be two measurable partitions of a measure space* $(\mathcal{K},\mu)$*, we denote*

(*a*) $d(\mathcal{P})=(\mu(P_1),\ldots,\mu(P_k)))$

(*b*) $d(\mathcal{P},\mathcal{Q})=\sum_{i=1}^{k}|\mu(P_i)-\mu(\mathcal{Q}_i)|$

e3.4 (*c*) if F is a measurable subset of $\mathcal{K}$ we denote by $\mathcal{P}/F$ (3.4)

the partition of F defined as $\mathcal{P}/F\stackrel{def}{=}(P_1\cap F,\ldots,P_k\cap F)$; and

$d(\mathcal{P}/F)\stackrel{def}{=}(\frac{\mu(P_1\cap F)}{\mu(F)},\ldots,\frac{\mu(P_k\cap F)}{\mu(F)})$.

Another basic tool in these lectures will be the following Rokhlin-Kakutani theorem (RK)

VIII. Theorem: *Let* $(\mathcal{K},\mu,T)$ *be an ergodic non atomic dynamical system. Given*

$\varepsilon > 0, N > 0$ there exists a measurable set $F \subset \mathcal{K}$ such that the sets $F, TF, \ldots, T^{N-1}F$ are pairwise disjoint and

e3.5
$$\mu(\cup_{i=0}^{N-1} T^i F) > 1 - \varepsilon \tag{3.5}$$

Furthermore if $\mathcal{Q} = (\mathcal{Q}_1, \ldots, \mathcal{Q}_n)$ is a given partition of $\mathcal{K}$ the set F could be so chosen that

e3.6
$$d(\mathcal{Q}/F) \;=\; d(\mathcal{Q}) \tag{3.6}$$

i.e. F could be such to be "divided" by $\mathcal{Q}$ in the same way in which $\mathcal{Q}$ "divides" the whole space.

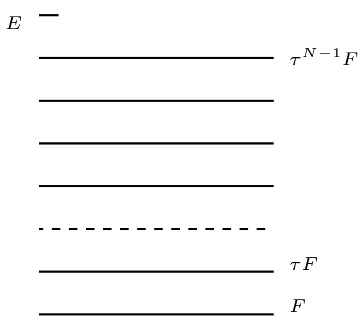

Fig. 1 The interval F, base of the tower, should be imagined divided into smaller intervals representing the sets $F \cap E\binom{0}{s_0}$; the interval E represents the points left out. The map T should be imagined to act by shifting upwards the points of each level with the exception of the highest level: the action of T on the latter will send its points part in the residual set E and the rest somewhere in the lowest level.

We shall refer to the objects defined by the RK theorem as the "*RK tower*", or "*RK-stack*" of $(\mathcal{K}, \mu.T)$ with base F, partition $\mathcal{Q}$, height N and approximation ε.

The reason for the name "tower" is that the sets $F, TF, \ldots, T^{N-1}F$ can be represented as N horizontal segments of length $\mu(F)$ on which T acts by shifting a point upwards to the next line along the vertical direction: in this picture the measure μ is represented as the Lebesgue measure on the segments; the set of segments plus a little extra piece E of length $(1 - N\mu(F)) < \varepsilon$ can represent the space $(\mathcal{K}, \mu)$: note, however, that the action of T on $T^{N-1}F$ or on E is not defined in the picture.

Lecture IV

In this lecture we consider two processes $(T, \mathcal{P})$, $(T', \mathcal{P}')$ extracted from two ergodic non atomic dynamical systems (but not necessarily isomorphic to them) and we *try to* "copy" $(T, \mathcal{P})$ into $(T', \mathcal{P}')$, *i.e.* we try to find a partition $\overline{\mathcal{P}} \in \vee_{-\infty}^{+\infty} T^{1^i} \mathcal{P}^1$ such that the process $(T', \overline{\mathcal{P}}')$ is a copy of $(T, \mathcal{P})$, i.e. it has the same joint distributions

e4.1
$$d(\bigvee_{i=0}^{N} T^{-i}\mathcal{P}) = d(\bigvee_{i=0}^{N} T'^{-i}\overline{\mathcal{P}}) \qquad \forall N \geq 0 \tag{4.1}$$

Assume, to begin with, that $(\mathcal{P}, T)$ and $(\mathcal{P}', T')$ have the same entropy s and that $s > 0$.

Let $\widetilde{\mathcal{P}}'$ be a partition measurable with respect to $\vee_{-\infty}^{+\infty}T'^{-i}\mathcal{P}'$ and assume that

e4.2
$$\mathcal{S}(\widetilde{\mathcal{P}}',T')-\mathcal{S}(\mathcal{P}',T')=-\delta, \qquad s>\delta>0 \tag{4.2}$$

(such a partition exists because of the continuity property (6) of theorem 2, Lecture III). The value of δ, will be chosen later.

Given $\varepsilon>0, u>0$ assume that $\sqrt{\varepsilon}<S(\mathcal{P}',T')$ and fix N very large (see later) and construct two RK towers with parameters ε, N associated, respectively, with $(\mathcal{P},T)$ and $(\widetilde{\mathcal{P}}',T')$ and with respective bases F,F' and partitions $\mathcal{Q}=\vee_{i=0}^{N-1}T^{-i}\mathcal{P}, \widetilde{\mathcal{Q}}'=\vee_{i=0}^{N-1}T'^{-i}\widetilde{\mathcal{P}}'$.

Assume that N is larger than the $N_{\varepsilon,u}$ needed for the conclusions of the Shannon-McMillan theorem to apply both to $(T,\mathcal{P})$ and $(T',\widetilde{\mathcal{P}}')$.

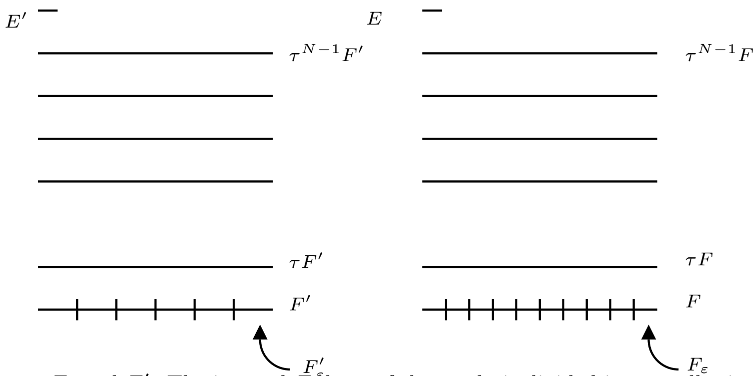

Fig. 2 The sets F_ε and F'_ε. The interval F, base of the stack, is divided into smaller intervals representing the sets $F\cap E\binom{0\ldots N-1}{\sigma_0\ldots\sigma_{N-1}}$ with $\sigma_0\ldots,\sigma_{N-1}$ chosen in the large frequency collection $\mathcal{C}^2_\varepsilon(N)$, except the rigthmost interval, denoted F_ε, which represents the intersection of F with $\cup E\binom{0\ldots N-1}{\sigma_0\ldots\sigma_{N-1}}$ where the union is over the cylinders with $\sigma_0,\ldots,\sigma_{N-1}$ in the collection $\mathcal{C}^2_\varepsilon(N)$ of rare strings.

For the purposes of better visualization we can assume that the atoms of $\mathcal{Q},\widetilde{\mathcal{Q}}'$ which belong to the classes $\mathcal{C}^2_N,{\mathcal{C}'}^2_N$ respectively, of theorem 4 lecture II, intersect F,F' in a sequence of adjacent intervals whose extremes are represented by tick-marks in fig. 2. The atoms in $\mathcal{C}^1_N,{\mathcal{C}'}^1_N$ intersect F,F' in a little interval $F_\varepsilon,F_\varepsilon$, drawn as the rightmost interval in the RK towers bases F,F' (respectively with $\mu(F_\varepsilon)\le\varepsilon\mu(F),\mu(F'_\varepsilon)\le\varepsilon\mu(F')$). In fig. 3 the dotted vertical lines are drawn to help visualize where the image of an inteval of the base is mapped, upon application of T or T' or their powers up to $n-1$.

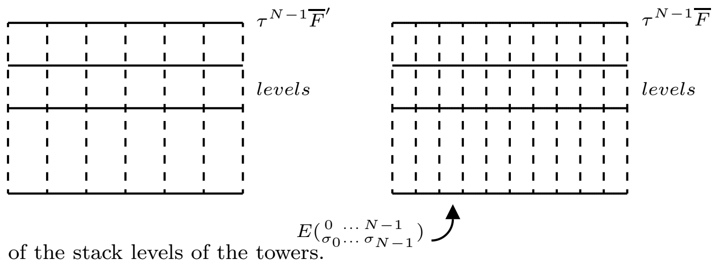

Fig. 3 Illustration of the stack levels of the towers.

The vertical dotted lines symbolize the points where the boundaries of the atoms in C^2_N,C'^2_N cut into smaller sets the sets $\overline{F}'=F'/F_\varepsilon$ and $\overline{F}=F/F_\varepsilon$ and their respective images under $T',\ldots,T'^{N-1}$ or $T,\ldots,T^{N-1}$: the smaller sets are represented by little segments which we call levels.

The levels can be used to reconstruct out of the above pictures the partitions $\widetilde{\mathcal{P}}'$ or $\mathcal{P}$ on the set $G' = \cup_{i=0}^{N-1} T'^i \overline{F}'$ or $G = \cup_{i=0}^{N-1} T^i F$ (i.e. almost perfectly since the sets G' or G fill up almost completely their respective spaces: more precisely the measure of their complements is bounded by 2ε): in fact if the atom of $\mathcal{Q}/\overline{F}$ at the base of a column is $(\cap_{i=0}^{N-1} T^{-i} P_{x_i}) \cap \overline{F}$ this means that this 0-th level (i.e. the segment $\cap_{i=0}^{N-1} T^{-i} P_{x_i} \cap \overline{F}$) is in P_{x_0} and the 1-th level above it is in P_{x_1}, etc.: so if we associate with each level the appropriate name we see that we reconstruct the set $P_\ell \cap G, \ell = 1, 2, \ldots, k$; similar argument holds for $\widetilde{P'}_\ell \cap G'$.

Note that in the picture we have drawn many more columns in the RK tower with base F than in the RK tower with base F': this is because of the Shannon-McMillan theorem and the fact that $S(\widetilde{\mathcal{P}}', T') < S(\mathcal{P}', T') = S(\mathcal{P}, T)$, and because we have implicitly assumed $2\varepsilon < \delta$ and N very large (this could have been achieved by choosing at the beginning $\delta = \sqrt{\varepsilon}$).

It is therefore possible to associate with each atom of $\widetilde{\mathcal{Q}}'/\overline{F}'$ an atom of $\mathcal{Q}/\overline{F}$ in such a way that different atoms of $\mathcal{Q}'/\overline{F}'$ have different correspondents in $\mathcal{Q}/\overline{F}$.

Each atom $\overline{F} \cap \bigcap_{i=0}^{N-1} T^{-i} P_{x_i} \in \mathcal{Q}/\overline{F}$ is determined by its "name" i.e. by the string $(x_0, \ldots, x_{N-1})$. We can use the name of the atom $q \in \mathcal{Q}/\overline{F}$ associated with an atom $q' \in \widetilde{\mathcal{Q}}'/\overline{F}$ to relabel the levels above q': *the i-th level $T'^i q'$ obtains name x_i if $(x_0, \ldots, x_{N-1})$ is the string associated with q.* If we collect together levels with the same new labels we construct k sets (if $\mathcal{P} = (P_1, \ldots, P_k)$) which we call $(\overline{P}'_1, \ldots, \overline{P}'_k)$ which partition G': we extend this partition, arbitrarily, outside G' and obtain a partition $\overline{\mathcal{P}}' = (\overline{P}'_1, \ldots, \overline{P}'_k)$ of the space $\mathcal{K}'$.

The new partition, because of the property 3) of the Shannon-McMillan theorem (theorem 4 of lecture II), is such that

e4.3
$$d(\vee_{i=0}^{N} T'^{-i} \overline{\mathcal{P}}', \vee_{i=0}^{N} T^{-i} \mathcal{P}) \tag{4.3}$$

is infinitesimal with $\delta, \varepsilon, 1/N$.

The partition $\overline{\mathcal{P}}'$ has also the property of having an entropy $S(\overline{\mathcal{P}}', T)$ such that $|S(\overline{\mathcal{P}}', T') - S(\mathcal{P}, T)|$ is infinitesimal as $\varepsilon \to 0, N \to \infty$.

The reason is connected with the fact that we have associated different strings to different atoms of $\widetilde{\mathcal{Q}}'/\overline{F}'$: this implies that the partitions of G' into the levels is refined by $\vee_{-N}^{N} T'^i (\overline{\mathcal{P}}' \vee \mathcal{F}')$ where $\mathcal{F}'$ is the partition of $\mathcal{K}'$ into $\overline{F}'$ and $\mathcal{K}'/F'$; the reader should look for the simple geometrical meaning of this statement.

Therefore the partition $\widetilde{\mathcal{Q}}'/G'$ is also refined by $\vee_{i=-N}^{N} T'^i (\overline{\mathcal{P}}' \vee \mathcal{F}')$ hence $\mathcal{P}'/G'$ is refined by $\vee_{i=-N}^{N} T'^i (\overline{\mathcal{P}}' \vee \mathcal{F}')$.

Since the complement of G' has measure $\le 2\varepsilon$ this means that

e4.4
$$\widetilde{\mathcal{P}}' \le_{2\varepsilon} \vee_{i=-N}^{N} T'^i (\overline{\mathcal{P}}' \vee \mathcal{F}) \tag{4.4}$$

where the symbol $\le_\varepsilon$ means that, by suitably collecting atoms of $\vee_{-N}^{N} T'^i (\overline{\mathcal{P}}' \vee \mathcal{F})$, it is possible to construct a partition $\widetilde{\mathcal{P}}'_\varepsilon \le \vee_{-N}^{N} T'^i (\overline{\mathcal{P}}' \vee \mathcal{F})$ refined by the partition $\vee_{i=-N}^{N} T'^i (\overline{\mathcal{P}}' \vee \mathcal{F}')$) such that $D(\widetilde{\mathcal{P}}', \widetilde{\mathcal{P}}'_\varepsilon) < 2\varepsilon$.

Therefore $S(\widetilde{\mathcal{P}}', T')$ is very close (because of item (6) in theorem VII, lecture III) to e4.5 $$\mathcal{S}(\widetilde{\mathcal{P}'}_\varepsilon,T') \le \mathcal{S}(\vee_{i=-N}^{N} T'^{i}(\overline{\mathcal{P}}' \vee \mathcal{F}),T') = \mathcal{S}(\overline{P}' \vee \mathcal{F}',T') \le \mathcal{S}(\overline{P}',T') + \mathcal{S}(\mathcal{F}), \quad (4.5)$$

Since $\mathcal{S}(\overline{P}',T) \le \mathcal{S}(\mathcal{P}',T) \le \mathcal{S}(\mathcal{P},T)$ this means that $|\mathcal{S}(\overline{P}',T') - S(\mathcal{P},T)|$ is infinitesimal with δ and ε (notice that $\mathcal{S}(\mathcal{F}) \to 0$ as $\varepsilon \to 0$).

So we have proved the following copying lemma

1. Lemma: *Given $\varepsilon > 0$, $u > 0$ and two ergodic non atomic processes with equal positive entropy, $(\mathcal{P},T)$ and $(\mathcal{P}',T')$, it is possible to find a partition $\overline{\mathcal{P}}'_{u,\varepsilon}$ of $\mathcal{K}'$ such that the processes $(\mathcal{P},T)$ and $(\overline{\mathcal{P}}'_{u,\varepsilon},T')$ have the properties:*

e4.6 $$d(\vee_{i=0}^{u} T^{-i}\mathcal{P}, \vee_{i=0}^{u} T'^{-i}\overline{\mathcal{P}}'_{u,\varepsilon}) < \varepsilon, \qquad |S(T,\mathcal{P}) - S(T',\overline{\mathcal{P}}'_{u,\varepsilon})| < \varepsilon \quad (4.6)$$

At this point it is natural to let $u \to 0$, $\varepsilon \to 0$. However the space of the partitions of $\mathcal{K}'$ with k atoms is not compact in any useful sense. So the risk is high that $\overline{\mathcal{P}}'_{u,\varepsilon}$ wanders without limit points in the space of the partitions.

We are stuck unless we make some further assumptions on the process $(T,\mathcal{P})$ that we are trying to copy: at this point the above lemma represents the best we can do without further assumptions.

In the next lecture we shall try to understand which is the property to impose on $(T,\mathcal{P})$ in order to be able to copy it in $(T',\mathcal{P}')$ by trying to improve the partition $\overline{\mathcal{P}}'_{u\,,\varepsilon}$ keeping track of how much the new partition differs from $\overline{\mathcal{P}}'_{u,\varepsilon}$.

Lecture V

Since the discussion of this lecture is essentially heuristic let us make the simplifying assumption that $\mathcal{S}(\overline{\mathcal{P}}'_{u,\varepsilon,T'}) < \mathcal{S}(\mathcal{P}',T')$ (a priori we might have equality).

Fix u',ε' and ε' will be much smaller than $\delta' = \mathcal{S}(\mathcal{P}',T') - \mathcal{S}(\overline{\mathcal{P}}'_{k,\varepsilon}T')$

Let $\mathcal{R}$ be a refinement of $\overline{\mathcal{P}}_{u,\varepsilon} \equiv \overline{\mathcal{P}}'$ (here after we suppress the two lower indices) such that $\mathcal{S}(\mathcal{R},T') < \mathcal{S}(\mathcal{P}',T') \equiv \mathcal{S}'$ and $\mathcal{S}(\mathcal{P}',T') - \mathcal{S}(\mathcal{R},T') = \mathcal{S}''$ very small compared to $\mathcal{S}'$; and assume that ε' is also much smaller than $\mathcal{S}''$ (i.e. the ratios $\frac{\mathcal{S}''}{\mathcal{S}'}, \frac{\varepsilon'}{\mathcal{S}''}$ are very small: see later).

We choose N' so large that $\frac{u'}{N'}$ is very small and Shannon-McMillan theorem with parameters ε', u' applies to the partition $\mathcal{P}$ with the transformation T and to the partitions $\mathcal{R}, P'$ with the transformation T'.

Construct the RK for the dynamical systems $(\mathcal{P},T)$ $(\mathcal{P}',T')$ and with bases F, F', parameters ε', N' and with partitions $\mathcal{Q} = \vee_0^{n-1} T^{-i}\mathcal{P}$ and $\vee_0^{N-1} T^{1-i}\mathcal{R}$.

We proceed exactly as before and construct the sets $\overline{F}, \overline{F}'$ which are partitioned by the "good" atoms of $\mathcal{Q}, \mathcal{Q}'' = \vee_{i=0}^{N-1} T'^{-i}\overline{\mathcal{P}}'/F'$.

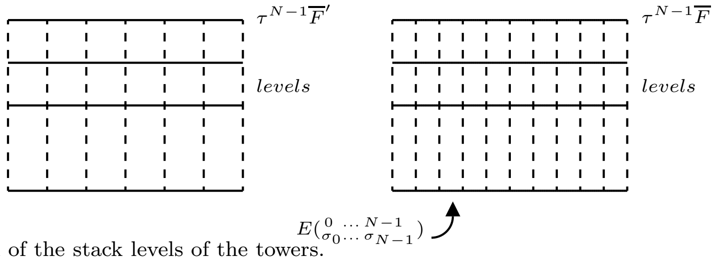

Fig. 3 Illustration of the stack levels of the towers.

The partitioning of $F' \cup T'F' \ldots \cup T'^{N-1}F'$ associated with the partitioning of F' into "good" atoms of $\mathcal{Q}''$ is denoted by the vertical continuous lines.

Since $\mathcal{R}$ refines $\overline{\mathcal{P}}$ the partitioning of $G' = \cup_{i=0}^{N-1} T'^{i}\overline{F}$ will be denoted by dotted lines and no continuous line can be between two dotted lines associated with a single atom of $\mathcal{Q}'$ (some dotted lines, however, may be superposed to continuous lines).

We denote the "bad" atoms of $\mathcal{Q}'$ which are contained in good atoms of $\mathcal{Q}''$ by a bold face segment, as in the picture: notice that the union of the bold face segments in F' and of their images under $T', \ldots, T'^{N-1}$ has measure $\le 2\varepsilon'$.

We could now construct $\overline{\mathcal{P}}'_{u',\varepsilon}$ by simply associating strings of "good" $\mathcal{Q}$-names to the "good" atoms of $\mathcal{Q}''$ and proceeding s in the lecture IV to construct $\overline{\mathcal{P}}_{u,\varepsilon}$.

The partition so obtained would however be quite unrelated to the old $\overline{\mathcal{P}}'$.

There is, however, a possibility which would allow us to have a $\overline{\mathcal{P}}'_{u',\varepsilon'}$, very close to $\overline{\mathcal{P}}'_{u,\varepsilon}$. Suppose that there existed a measure preserving mapping between F and F' (thought as probability spaces with the measure μ and μ' (respectively) suitably renormalized) such that corresponding points have names with respect to the partitions $\vee_{i=0}^{N'} T^{-i}\mathcal{P}$ and $\vee_{i=0}^{N'} T'^{-i}\overline{\mathcal{P}}'_{u,\varepsilon}$ which coincide for all but $\theta(\varepsilon,\mu)N'$ indices where $\theta(\varepsilon,\mu) \xrightarrow[\varepsilon\to 0]{} 0$, with the possible exception of a set of "bad" points with measure $\le \theta(\varepsilon,u)$.

Then we could use this mapping ψ to wisely assign new names to the good atoms of $\mathcal{Q}'/F'$: precisely if a "good" atom q' of $\mathcal{Q}'/F'$ contains the image of a good point x which was in a good atom q of $\mathcal{Q}/F$ then we give to q' the name of q. Of course there are some ambiguities: i.e. the same q' may contain images of two good points which belong to two different atoms. These ambiguities may be resolved by making an arbitrary choice provided care s taken to make sure that different "good" atoms of $\mathcal{Q}'/F'$ receive different names (here of course enters the fact that if N' is large and the ratios $\frac{\varepsilon'}{S''}, \frac{\mathcal{S}'}{S''}$ are small the number of "good" atoms of $\mathcal{Q}'/F'$ greatly exceeds that of the good atoms of $\mathcal{Q}''/F'$ while on the other hand, their measure are arranged in the opposite scale. We do not enter into the details of this construction because they are of technical nature and not too difficult and, in my opinion, they are not the central point of the proof.

It is now clear that the new partition $\overline{\mathcal{P}}'_{u',\varepsilon'}$, so constructed will, on each column with base a "good ‘' segment of $\mathcal{Q}'/F'$, differ from $\overline{\mathcal{P}}'_{u,\varepsilon}$ only on a set of relative measure $\theta(\varepsilon,\mu)$ with the possible exception of a set of columns whose total measure adds to less than $\theta(\varepsilon,\mu) + 2\varepsilon'$. This means that

e5.1
$$D(\overline{\mathcal{P}}'_{u,\varepsilon}, \overline{\mathcal{P}}'_{u',\varepsilon'}) \le 2\theta(\varepsilon,\mu) + 3\varepsilon' \tag{5.1}$$

Hence the new process $(T', \overline{\mathcal{P}}_{u',\varepsilon'})$ will be a far better approximation to $(T, \mathcal{P})$ and, at the same time, it will be only $2\theta(\varepsilon,\mu) + 3\varepsilon'$ far from $\overline{\mathcal{P}}'_{u,\varepsilon}$.

Since u', ε' are arbitrary it is clear that we can iterate the procedure in such a way that the changes in $\overline{\mathcal{P}}'_{u,\varepsilon}$ from one approximation to the next are so small that there is a limit partition $\widehat{\mathcal{P}}'$ such that $(T', \widehat{\mathcal{P}})$ is an exact copy of $(T, \mathcal{P})$.

Note that to perform the above construction one does not really need a measure preserving map of F onto F': it would be enough to have an isomorphism of the measure spaces $(F, \frac{\mu}{\mu(F)})$ onto $(F', \frac{\mu}{\mu'(F')})$ such that: if $a \in \vee_{i=0}^{N-1} T^{-i}\mathcal{P}/F$ then its image $\psi(a)$ consists in points which have $\mathcal{Q}'$-name differing from the $\mathcal{Q}$-name of a by at most $\theta(\ell,\mu)N'$ entries aside from a few exceptional points which add up to a set of measure $\le \theta(\ell,\mu)$.

This remark is important because it shows that the existence or not of the map ψ is only a property of the numbers $d(\vee_{i=0}^{N'-1} T'^{-i}\overline{\mathcal{P}}'/F') = d(\vee_{i=0}^{N'-1} T^{-i}\overline{\mathcal{P}}') = d(\mathcal{Q}')$ and $d(\vee_{i=0}^{N'-1} T^{-i}\mathcal{P}/F) = d(\vee_{i=0}^{N-1} T^{-i}\mathcal{P}) = d(\mathcal{Q})$: i.e. it does not depend upon F, F' but only upon $(\overline{\mathcal{P}}', T'), (\mathcal{P}, T)$.

The conclusion of the above semi-heuristic discussion is that in order to be really able to perform the above construction (aside from the mentioned technicalities) we need to be able to infer from the information (which can in the particular cases be extracted from the lemma 1 of Lecture IV) that if a process $(T', \overline{\mathcal{P}}')$ imitates $(T, \mathcal{P})$ in the sense that for some $n_\varepsilon, \delta_\varepsilon$, (given $\varepsilon > 0$):

e5.2
$$d\,(\vee_{i=0}^{n_\varepsilon} T^{-i}\mathcal{P}, \vee_{i=0}^{n_\varepsilon} T'^{-i}\overline{\mathcal{P}}) < \delta_\varepsilon, \qquad |S(T', \overline{\mathcal{P}}') - S(T, \mathcal{P})| < \delta_\varepsilon \tag{5.2}$$

then, for every fixed N, it must be possible to construct a isomorphism ψ between $(\mathcal{K}, \mu)$ and $(\mathcal{K}', \mu')$ (see theorem 3) such that the points of the images $\psi(a))$ of atoms $a \in \vee_{i=0}^{N} T^{-i}\mathcal{P}$ have names with respect to $\vee_{i=0}^{N} T'^{-1}\overline{\mathcal{P}}$ which agree for all but a set of εN indices aside from a set of possible exceptions which has total measure $\le \varepsilon$.

This property however does not at all follow in general.

Processes $(T, \mathcal{P})$ for which the above property holds can be naturally called "*finitely determined*".

This notion is particularly interesting since it can be proven, without too much work, that finitely determined processes exists and in fact every Bernoulli scheme is finitely determined.

Ornstein has been able to show much more: namely a process is isomorphic to a Bernoulli scheme if and only if it is finitely determined.

It is clear how the above construction, if completely performed in its technical details, would then lead to the conclusion that it is possible to copy any finitely determined process $(T, \mathcal{P})$ into any ergodic non atomic process $(T', \mathcal{P}')$ with the same entropy.

Together with the fact that Bernoulli schemes are finitely determined, this implies the weak isomorphism of two Bernoulli schemes with the same entropy.

I stop here these lectures after giving only one references: [1], for further reading; in the reference all the details missing in these lectures and many more insights into the theory of the isomorphism between ergodic processes can be found.

Finally let me thank the participants in the school for correcting several mistakes present in the draft of these notes.

References

[1] D. Ornstein: *Ergodic Theory and Randomness*, Yale Univ. Press, 1974.

Problem 1: (Rohlin's stack), Under the hypothesis of *theorem IV* condider a Borel partition $\mathcal{Q} = (Q_0, \ldots, Q_k)$ of $\{0, \ldots, n\}^{\mathbb{Z}}$. Show that the first statement of the theorem implies the second if the system is mixing: *i.e.* it is possible, given N and $\varepsilon > 0$, to find F so that $\mu(Q_i \bigcap F)/\mu(F) = \mu(Q_p)$ for all $i = 0, \ldots, k$, and $\tau^i F \cap \tau^j F = \emptyset$, $\forall\, 0 \le i \ne j \le N$. (*Hint:* Let F_0 be the set whose existence, in correspondence of the given N and ε, is assured by the first statement in *theorem IV* and which satisfies the properties described therein. Set $F_t = \tau^t F_0$, $t \in \mathbb{Z}$, and use the mixing property (assumed to hold for μ) to infer that

e5.3 $$\lim_{t\to\infty} \sum_{i=1}^{k} \Big| \frac{\mu(Q_i \bigcap F_t)}{\mu(F_t)} - \mu(Q_i) \Big| = \lim_{t\to\infty} \eta_t = 0 \tag{5.3}$$

If one chooses $\eta_t \ll \mu(F_t) = \mu(F_0)$ (by the τ-invarience of μ) it is clear that, being $(\{0, \ldots, n\}^{\mathbb{Z}}, \mu)$ isomorphic mod 0 to the Lebesgue measure on $[0, 1]$ (*cf. theorem II*) and regarding in this way $Q_0, \ldots, Q_k$ and F_t as sets of $[0, 1]$ it is possible to take out of F_t a set $\Delta \subset F_t$ of points having small measure with respect to $\eta_t \ll \mu(F_0)$ to obtain that $\mu(Q_i \bigcap (F_t \setminus \Delta)) = \mu(Q_i)\mu(F_t \setminus \Delta)$ without deteriorating the bound on the measure of $\bigcup_{i=0}^{N-1} \tau^i (F_t \setminus \Delta))$, *i.e.* keeping it larger than $1 - \varepsilon$).

Remark: *Rohlin's stack theorem* does not require the hypothesis of mixing: ergodicity of μ suffices; however the proof is, in the latter cases, somewhat more elaborate.

Problem 2: (Sinai' theorem) Let $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, \mu)$ and $(\{0, \ldots, n'\}^{\mathbb{Z}}, \tau, \mu')$ be two mixing shifts with $s(\mu', \tau) > s(\mu, \tau)$. Consider the partitions $\mathcal{P}$ and $\mathcal{P}'$ of $\{0, \ldots, n\}^{\mathbb{Z}}$ and of $\{0, \ldots, n'\}^{\mathbb{Z}}$ into the cylinders with base 0 (*i.e.* $\mathcal{P} = \{C^0_0, C^0_1, \ldots, C^0_n\}$ and $\mathcal{P}' = \{C'^0_0, \ldots, C'^0_{n'}\}$ respectively).

Given $\varepsilon > 0$ and u integer > 0 consider the partitions $\mathcal{P}_N = \vee_0^{N-1} \tau^i \mathcal{P}$ and $\mathcal{P}'_N = \vee_0^{N-1} \tau^i \mathcal{P}'$ and choose $N > N(u, \varepsilon)$, where $N(u, \varepsilon)$ is such that for $N > N(u, \varepsilon)$ the approximability in distribution and entropy hold. Let $F \subset \{0, \ldots, n\}^{\mathbb{Z}}$ and $F' \subset \{0, \ldots, n'\}^{\mathbb{Z}}$ be two sets for which (*cf.* problem 1)

$$\frac{\mu(Q \bigcap F)}{\mu(F)} = \mu(Q), \text{ for all } Q \in \mathcal{P}_N \, ; \; \frac{\mu'(Q' \bigcap F')}{\mu'(F')} = \mu'(Q') \text{ for all } Q' \in \mathcal{P}'_N$$

and simultaneously $\tau^i F \bigcap \tau^j F = \emptyset = \tau^i F' \bigcap \tau^j F'$, for all $i \ne j$, $i, j = 0, \ldots, N-1$. Represent F and F' as two intervals (see Fig.2), and represent as intervals also the sets

$$\tau F, \ldots, \tau^{N-1} F, \qquad \tau F', \ldots, \tau^{N-1} F'$$

$$E = \{0, \ldots, n\}^{\mathbb{Z}} / \bigcup_{i=0}^{N-1} \tau^i F, \qquad E' = \{0, \ldots, n'\}^{\mathbb{Z}} / \bigcup_{i=0}^{N-1} \tau^i F'.$$

Check that the action of τ is naturally represented as an upward translation except for its action on E and E' and on $\tau^{N-1} F$ and $\tau^{N-1} F'$ (where it acts differently and in a way which, in general, is not simply representable graphically): in this representation the measures μ and μ' are represented by the Lebesgue measure on the several intervals whose lengths, in every stack, add up to 1.

Problem 3: In the situation of problem 2 draw as segments the several elements of the partition induced by $\mathcal{Q} \equiv \mathcal{P}_N$ on F (*cf.* Fig.2): $(Q \bigcap F)_{Q\in\mathcal{Q}}$; and represent as an interval also the set $F_\varepsilon = \bigcup F \bigcap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ where the union is over the choices of $(\sigma_0, \ldots, \sigma_{N-1})$ in the collection $\mathcal{C}_{2,\varepsilon,u}(N)$ of sets realizing the aproximation in entropy and distribution. Perform the same construction over the stack relative to F'. Check that $\mu(F_\varepsilon) \le \varepsilon\mu(F)$, $\mu'(F'_\varepsilon) \le \varepsilon\mu'(F')$.

Problem 4: In the situation of problems 1,2 set $\overline{F} = F/F_\varepsilon$ and $\overline{F}' = F'/F'_\varepsilon$: such sets are split into disjoint parts by the partitions $\overline{F} \bigcap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ and $\overline{F}' \bigcap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ with $\sigma_0, \ldots, \sigma_{N-1} \in \mathcal{C}_{1,\varepsilon,u}(N)$ and $\sigma'_0, \ldots, \sigma'_{N-1} \in \mathcal{C}'_{1,\varepsilon,u}(N)$ (*cf.* problem 2). Fix also $N \gg u$. We shall call "level" of the stack every image $\tau^j(\overline{F} \bigcap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}})$, with $0 \le j \le N-1$ and $\sigma_0, \ldots, \sigma_{N-1} \in \mathcal{C}_{1,\varepsilon,u}(N)$. Likewise we define the levels for the stack with base $\overline{F}'$. See Fig.3.

Remark that by the Shannon–McMillan theorem, if $2\varepsilon < s(\mu') - s(\mu)$ and N is large enough, the "columns of levels" with base F' are much more numerous of those with base F.

Put arbitrarily into correspondence every column with base on F with a different column with base on F' by assigning to the j–th level of a column with base on F the symbol σ'_j of the column with base on F' associated with it.

Collecting then the levels that, in this construction, come to have labels equal to $0, 1, \ldots, n'$ respectively, form a partition of $\bigcup_{j=0}^{N-1} \tau^j \overline{F}$ in $n'+1$ sets $\tilde{P}'_0, \ldots, \tilde{P}'_{n'}$: imagine extending such a partition to a partition $\tilde{\mathcal{P}}' = \{\tilde{P}'_0, \ldots, \tilde{P}'_n\}$ of the whole space $\Omega = \{0, \ldots, n\}^{\mathbb{Z}}$, arbitrarily.

Show that if $\mathcal{F}$ is the partition of $\{0, \ldots, n\}^{\mathbb{Z}}$ into $\overline{F}$ and $\{0, \ldots, n\}^{\mathbb{Z}}/\overline{F}$ we get that $\vee_{-N}^{N} \tau^i(\widetilde{\mathcal{P}}' \vee \mathcal{F})$ contains a partition $\widetilde{\mathcal{P}}$ with n elements formed by unions of its atoms and such that: $d(\mathcal{P}, \widetilde{\mathcal{P}}) < 2\varepsilon$. (*Hint:*The partition $\vee_{-N}^{N} \tau^i(\widetilde{\mathcal{P}}' \vee \mathcal{F})$ reduces on $\bigcup_{i=0}^{N-1} \tau^i \overline{F}$ to the partition into levels and, therefore, we reconstruct from it, by means of operations of unions of atoms, the partition $\mathcal{P}$ on $\bigcup_{i=0}^{N-1} \tau^i \overline{F}$, *etc.*)

Problem 5: Show that the partition $\widetilde{\mathcal{P}}$ built via the procedure illustrated in problems 1–3, is such that

$$\sum_{\sigma'_0\cdots\sigma'_{n-1}} |\mu'(C^{0\quad u-1}_{\sigma'_0\cdots\sigma'_{n-1}}) - \mu(P^{0\ldots u-1}_{\sigma'_0\cdots\sigma'_{n-1}})| \xrightarrow[N\to\infty,\varepsilon\to 0]{} 0$$

$$|s(\widetilde{\mathcal{P}}', \tau) - s(\mathcal{P}, \tau)| \xrightarrow[\varepsilon\to 0]{} 0$$

It is therefore possible to "represent a process with larger entropy into one of lower entropy and within a prefixed approximation" without losing in entropy more than a prefixed quantity beyond the obviously necessary loss $(s(\mu', \tau') - s(\mu, \tau))$. (*Hint:*Use the Shannon–McMillan theorem and the fact that if N is very large the strings (short because the length u is fixed) $\sigma'_0, \ldots, \sigma'_{u-1}$ appear with frequency almost equal to $\mu'(C^{0\ldots u-1}_{\sigma'_0\cdots\sigma'_{n-1}})$ in the sequences of $\mathcal{C}_{1,\varepsilon,u}(N)$.)

Problem 6: (*Copying a dynamical system into another*)
Deduce from the results of problem 4 that, under the same hypotheses of problem (2) but with $s(\mu', \tau) = s(\mu, \tau) = s$, then, given a positive integer u and given $\varepsilon > 0$, it is

possible to "copy" the first dynamical system into the second in the sense that it is possible to construct a partition $\overline{\mathcal{P}}$ of $\{0, \ldots, n'\}^{\mathbb{Z}}$ so that

$$\begin{aligned}&|s(\overline{\mathcal{P}}, \tau) - s| < \varepsilon\\ &\sum_{\sigma_0 \ldots \sigma_{n-1} \in \{0,\ldots,n\}^u} |\mu(C^{0\ldots u-1}_{\sigma_0 \ldots \sigma_{n-1}}) - \mu'(\overline{P}^{0\ldots u-1}_{\sigma_0 \ldots \sigma_{u-1}})| < \varepsilon\end{aligned}$$

(*Hint:*Find in $\{0, \ldots, n'\}^{\mathbb{Z}}$ a partition $\tilde{\tilde{\mathcal{P}}}'$ such that $s(\tilde{\tilde{\mathcal{P}}}', \tau) = s - 2\varepsilon$. Apply then again the construction of the Problems 2,3,4 replacing $\mathcal{P}'$ with $\tilde{\tilde{\mathcal{P}}}'$ *etc*).