CHAPTER III

Entropy and complexity

§3.1 Complexity of motions and entropy

Continuing the analysis of general structural properties of motions of an invertible discrete dynamical system (Ω, S) we shall now discuss the foundations of the notion of complexity of motions on Ω and of its theory.

Let $\mathcal{P}$ be a partition of Ω: the complexity of a motion on Ω as observed on $\mathcal{P}$ will be defined in terms of the complexity of its $(\mathcal{P}, S)$–history. Therefore we begin by discussing the notion of complexity of a sequence $\underline{\sigma}$.

N3.1.1 If $\widehat{\underline{\sigma}}$ is a sequence $\widehat{\underline{\sigma}} \in \{0, \ldots, n\}^{\mathbb{Z}}$ of symbols with defined frequencies[1] and if $N > 0$ we consider, as $(\sigma_0 \ldots \sigma_{N-1}) \in \{0, \ldots, n\}^N$ varies, the strings of history homologue to $\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix}$ that "appear" in $\widehat{\underline{\sigma}}$, *i.e.* which have strictly positive frequency of appearance. We shall set

e3.1.1
$$\eta_{\rm abs}(\widehat{\underline{\sigma}} \,|\, N) = \{\text{number of distinct } N\text{–length strings that appear in } \widehat{\underline{\sigma}}\}. \tag{3.1.1}$$

It would at first seem natural to identify the size of the *complexity* of $\widehat{\underline{\sigma}}$ with the number

e3.1.2
$$s_{\rm abs}(\widehat{\underline{\sigma}}) = \limsup_{N\to+\infty} N^{-1} \log \eta_{\rm abs}(\widehat{\underline{\sigma}}|N) = \lim_{N\to\infty} N^{-1} \log \eta_{\rm abs}(\widehat{\underline{\sigma}}|N), \tag{3.1.2}$$

[1] Cfr. definition (1.4.3).

where the logarithm and the factor N^{-1} are suggested by the expectation that $\eta_{\rm abs}(\widehat{\underline\sigma}|N)$ grows exponentially with N. The fact that

e3.1.3
$$\eta_{\rm abs}(\widehat{\underline\sigma}|N+M)\le\eta_{\rm abs}(\widehat{\underline\sigma}|N)\,\eta_{\rm abs}(\widehat{\underline\sigma}|M) \tag{3.1.3}$$

and $\eta_{\rm abs}(\widehat{\underline\sigma}|N)\le(n+1)^N$ imply the existence of the limit in (3.1.2). Indeed one has

$$f_k\equiv 2^{-k}\log\eta_{\rm abs}(\widehat{\underline\sigma}|2^k),\;\; f_k\le\log(n+1),\;\; f_{k+1}\le f_k,\;\; \text{for all}\, k>0,$$

so that $\lim_{k\to\infty} f_k$ exists. Moreover, given k and taking N large one can write $N=m\,2^k+r$ and obtain that

$$\frac1N\log\eta_{\rm abs}(\widehat{\underline\sigma}|N)\le f_k+\frac1N\log\eta_{\rm abs}(\widehat{\underline\sigma}|r),$$

so that eventually $N^{-1}\log\eta_{\rm abs}(\widehat{\underline\sigma}|N)\le f_k+\varepsilon$ for every ε. With a similar reasoning, writing $2^k=m\,N+r$ with k large, one shows that also the inequality

$$\frac1N\log\eta_{\rm abs}(\widehat{\underline\sigma}|N)\ge f_k-\frac1{2^k}\log\eta_{\rm abs}(\widehat{\underline\sigma}|r)$$

holds. This proves that the limit exists.

However, obviously, this is not the only possible definition of complexity of $\widehat{\underline\sigma}$: for example here we put on the same level strings of history $\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$ with very different frequency of appearance in $\widehat{\underline\sigma}$, $p\Big(\begin{matrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{matrix}\,|\,\widehat{\underline\sigma}\Big)$.

The following definition takes into some account the possible existence of very many history strings with low frequency of appearance in $\widehat{\underline\sigma}$.

Given $\varepsilon>0$ we shall consider all possible partitions of the strings of length N in two classes $\mathcal C_1$ and $\mathcal C_2$ such that

e3.1.4
$$\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal C_2} p\Big(\begin{matrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{matrix}\Big|\widehat{\underline\sigma}\Big)\le\varepsilon \tag{3.1.4}$$

setting

e3.1.5
$$\eta_\varepsilon(\widehat{\underline\sigma}|N)=\inf_{\mathcal C_1,\mathcal C_2}\{\text{number of elements of } \mathcal C_1\}, \tag{3.1.5}$$

and noting that $\eta_\varepsilon(\widehat{\underline\sigma}|N)$ does not decrease as ε decreases, at N fixed, we shall define

e3.1.6
$$s(\widehat{\underline\sigma})=\lim_{\varepsilon\to0}\limsup_{N\to\infty}N^{-1}\log\eta_\varepsilon(\widehat{\underline\sigma}|N). \tag{3.1.6}$$

Note the relation between (3.1.2) and (3.1.6) by writing $s_{\rm abs}(\widehat{\underline\sigma})$ as

e3.1.7
$$s_{\rm abs}(\widehat{\underline\sigma})=\limsup_{N\to\infty}\lim_{\varepsilon\to0}N^{-1}\log\eta_\varepsilon(\widehat{\underline\sigma}|N). \tag{3.1.7}$$

when $\underline{\sigma}$ has defined frequencies.

Having seen the above two possible notions of complexity several others come to mind, depending on whether one wishes to judge of "minor importance" certain strings of history relative to the "importance" of others.

Here is a rather general definition. Let $\underline{V} = \{V_N\}_{N=1}^{\infty}$ be a sequence of functions $(\sigma_0 \ldots \sigma_{N-1}) \to V_N(\sigma_0 \ldots \sigma_{N-1})$ defined respectively on $\{0, \ldots, n\}^N$, $N = 1, 2, \ldots$. Given an infinite sequence $\widehat{\underline{\sigma}} \in \{0, \ldots, n\}^{\mathbb{Z}}$ with defined frequencies, consider the subdivisions $\mathcal{C}_1$, $\mathcal{C}_2$ of $\{0, \ldots, n\}^N$ in two classes such that

e3.1.8
$$\sum_{\sigma_0 \ldots \sigma_{N-1} \in \mathcal{C}_2} p\Big(\begin{matrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{matrix} \Big| \widehat{\underline{\sigma}} \Big) \le \varepsilon \tag{3.1.8}$$

and set

e3.1.9
$$\eta_\varepsilon(\widehat{\underline{\sigma}} \,|\, N; \underline{V}) = \inf_{\mathcal{C}_1, \mathcal{C}_2} \sum_{\sigma_0 \ldots \sigma_{N-1} \in \mathcal{C}_1} e^{-V_N(\sigma_0 \ldots \sigma_{N-1})}. \tag{3.1.9}$$

Clearly $\eta_\varepsilon(\widehat{\underline{\sigma}} \,|\, N; \underline{0}) = \eta_\varepsilon(\widehat{\underline{\sigma}} \,|\, N)$. We shall set

e3.1.10
$$s(\widehat{\underline{\sigma}} \,|\, \underline{V}) = \lim_{\varepsilon \to 0} \limsup_{N \to \infty} N^{-1} \log \eta_\varepsilon(\widehat{\underline{\sigma}} \,|\, N; \underline{V}) \tag{3.1.10}$$

N3.1.2 and call this quantity the *complexity with weight* $e^{-\underline{V}}$.[2]

We summarize the above discussion in the following definition, where we shall make use of the symbols introduced so far without discussing them again.

D3.1.1 **(3.1.1) Definition:** (Complexity and entropy of a sequence)
If $\widehat{\underline{\sigma}}$ is a sequence in $\{0, \ldots, n\}^{\mathbb{Z}}$ with defined frequencies and with distri-
N3.1.3 *bution $\underline{p}$* [3] *and if $\underline{V} = \{V_N\}_{N \ge 1}$ is a sequence of functions on $\{0, \ldots, n\}^N$, $N = 1 \ldots$, respectively, one defines the* complexity of $\widehat{\underline{\sigma}}$ with weight $\underline{V}$ *the quantity (3.1.10). If $V_N = 0$ such a quantity takes the name of* entropy *and it is given by (3.1.6).*

As a first application we evaluate the complexity of the motions of analytically integrable systems, *i.e.* of motions associated with the rotations of a torus $\mathbb{T}^r$ observed on an analytically regular partition $\mathcal{P} = \{P_0, \ldots, P_n\}$, see definition (1.4.1).

P3.1.1 **(3.1.1) Proposition:** *If $S : \underline{\varphi} \to \underline{\varphi} + \underline{\omega}$ mod 2π, $\underline{\omega} \in \mathbb{R}^r$, is a rotation of $\mathbb{T}^r$ and if $\mathcal{P}$ is an analytically regular partition of $\mathbb{T}^r$ one has*

e3.1.11
$$s(\underline{\sigma}(\underline{\varphi})) = 0 \qquad \forall\, \underline{\varphi} \in \mathbb{T}^r. \tag{3.1.11}$$

One says, in a colorful language, that "the entropy of quasi-periodic motions is zero".

---

[2] There is nothing "mysterious" in the exponential: it is just a way to define the weight so that it is automatically non negative.

[3] cf. remark (2) to definition (2.3.2)

Remarks: (1) As in the case of proposition (2.1.1) if $(\underline{\omega}, 2\pi)$ are rationally independent one can take the atoms of $\mathcal{P}$ as Riemann measurable sets. The higher regularity is necessary to cover the general situation, see (2.1.2). See also problems [3.1.10] and [3.1.11] below to understand the essential role of the regularity hypothesis on the partition $\mathcal{P}$.

Proof: For simplicity we shall consider only the case in which $(\omega_1, \ldots, \omega_r, 2\pi)$ are $r+1$ rationally independent numbers.

In this case the distribution $\underline{p}$ of the history $\underline{\sigma}(\underline{\varphi})$ of $\underline{\varphi} \in \mathbb{T}^r$ is given by

e3.1.12
$$p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} = \int_{P^{0 \ldots N-1}_{\sigma_0 \ldots \sigma_{N-1}}} \frac{d\underline{\varphi}}{(2\pi)^r}; \qquad (3.1.12)$$

cf. proposition (2.1.2) and equation (2.1.9).

We start with some geometric considerations. By the regularity hypothesis on $\mathcal{P}$ the surface area $|\,\partial P_\sigma\,|$ of the elements $P_\sigma \in \mathcal{P}$ is finite and we can set

e3.1.13
$$2L \stackrel{def}{=} \sum_{\sigma=0}^{n} |\,\partial P_\sigma\,| < +\infty. \qquad (3.1.13)$$

The sum of the areas of the boundaries of the sets $P^{0 \ldots N-1}_{\sigma_0 \ldots \sigma_{N-1}}$ is, obviously, such that

e3.1.14
$$\begin{aligned} \sum_{\sigma_0 \ldots \sigma_{N-1}} |\,\partial P^{0 \ldots N-1}_{\sigma_0 \ldots \sigma_{N-1}}\,| &\le \sum_{k=0}^{N-1} \sum_{\sigma=0}^{n} |\,S^{-k} \partial P_\sigma\,| \\ &= \sum_{k=0}^{N-1} \sum_{\sigma=0}^{n} |\,\partial P_\sigma\,| = 2LN, \end{aligned} \qquad (3.1.14)$$

because the rotation S is rigid: therefore it does not alter lengths, areas or volumes.

The volume of $P^{0 \ldots N-1}_{\sigma_0 \ldots \sigma_{N-1}}$ defined by (3.1.12) can be bounded, if the diameter of $P^{0 \ldots N-1}_{\sigma_0 \ldots \sigma_{N-1}}$ is small enough with respect to 2π (*e.g.* $< \pi$), by [4]
N3.1.4

e3.1.15
$$p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \le \Gamma_r\, |\,\partial P^{0 \ldots N-1}_{\sigma_0 \ldots \sigma_{N-1}}\,|^{\,r/(r-1)}, \qquad (3.1.15)$$

where Γ_r is a suitable constant easily expressible in terms of the volume of the unit r–dimensional sphere.

In general it is not possible to construct an analytically regular partition (see definition (1.4.1)) of the torus $\mathbb{T}^r$ composed by sets with small diameter. To make use of (3.1.15) we shall suppose that all the sets of $\mathcal{P}$ but one are, since the beginning, so small that $\mathrm{diam}(P_\sigma) < \pi$. We will call the set with

---

[4] This is the isoperimetric inequality: in $\mathbb{R}^r$ it is valid for every set, independently on the size of its diameter; on $\mathbb{T}^r$ it is necessary that the set “does not wrap around the torus”. For sets with small enough diameter, it follows from the isoperimetric inequality in $\mathbb{R}^r$.

large diameter P_0. The general case can be treated analogously. In this situation we have that diam $(P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}) \ge \pi$ only if $\sigma_0 = \ldots = \sigma_N = 0$. In the following argument we can include the set $P^{0\ldots N-1}_{0\ldots 0}$ in the set $\mathcal{C}_1(N)$ for every N, without affecting the validity of the argument.

Let us define, given $\eta > 0$,

e3.1.16

$$\begin{aligned}\mathcal{C}_1(N) &= \Big\{\sigma_0,\ldots,\sigma_{N-1}\,\Big|\, p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} > e^{-N\eta}\Big\},\\ \mathcal{C}_2(N) &= \Big\{\sigma_0,\ldots,\sigma_{N-1}\,\Big|\, p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le e^{-N\eta}\Big\}.\end{aligned} \tag{3.1.16}$$

We get, for all $\gamma > 0$ and taking advantage of the well known idea behind Chebishev's inequality and of (3.1.15),

e3.1.17

$$\begin{aligned}&\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_2(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le\\ &\le \sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_2(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\left(\frac{e^{-N\eta}}{p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}}\right)^{\gamma} \le\\ &\le e^{-N\eta\gamma}\,\Gamma_r^{1-\gamma}\sum_{\sigma_0\ldots\sigma_{N-1}} |\,\partial P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}\,|^{(1-\gamma)r/(r-1)},\end{aligned} \tag{3.1.17}$$

and selecting $\gamma = r^{-1}$, so that $(1-\gamma)r/(r-1) \equiv 1$, equations (3.1.17) and (3.1.14) imply

e3.1.18

$$\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_2(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le 2LN\,\Gamma_r^{1-\gamma}\,e^{-N\eta\gamma}. \tag{3.1.18}$$

N3.1.5 On the other hand {number of elements in $\mathcal{C}_1(N)$} $\le e^{+N\eta} + 1$:[5] hence $s(\underline{\sigma}(\underline{\varphi})) \le \eta$ for all $\eta > 0$ (note that the $s(\underline{\sigma}(\underline{\varphi}))$, as it is defined in (3.1.6), is less than the quantity computed by setting $\varepsilon = e^{-N\eta}$, as we are doing). This means that $s(\underline{\sigma}(\underline{\varphi})) = 0$. ■

The preceding proof does not provide us with an optimal result, nor it is the simplest conceivable: under the same hypothesis we could easily deduce that even $s_{\rm abs}(\underline{\sigma}(\underline{\varphi})) = 0$. However the proof just given is more interesting because it can be easily extended to much more general situations: for instance to those contemplated in the following definition and proposition.

D3.1.2 **(3.1.2) Definition:** *Let (Ω, S) be an invertible dynamical system. Let μ_0 be a probability measure (not necessarily S–invariant) on Ω and let $\mathcal{P}$ be a μ_0–measurable partition of Ω.*

---

[5] The $+1$ is here to take into account the set $P^{0\ldots N-1}_{0\ldots N-1}$ which is included in $\mathcal{C}_1$.

(i) Denote by $\widehat{\mathcal{E}}(S,\mathcal{P},\mu_0)$ *the symbolic motions* $\underline{\widehat\sigma}\in\widehat\Omega$*, observed on* $\mathcal{P}$ *with well defined frequencies and with distribution that can be expressed by integrals with respect to* μ_0 *as*

e3.1.19
$$p\begin{pmatrix} j_1\ldots j_q \\ \sigma_1\ldots\sigma_q\end{pmatrix} \mid \underline{\widehat\sigma}\Big) = \int_{P^{j_1\ldots j_q}_{\sigma_1\ldots\sigma_q}} \rho(x)\mu_0(dx), \tag{3.1.19}$$

where ρ *and also* $\rho^{-1}\in L_1(\mu_0)$*.*

We shall call the points of $\widehat{\mathcal{E}}(S,\mathcal{P},\mu_0)$ symbolic motions which are absolutely continuous with respect to μ_0*.*

(ii) The points of Ω *whose* $(\mathcal{P},S)$*–histories are in* $\widehat{\mathcal{E}}(S,\mathcal{P},\mu_0)$ *will be denoted* $\mathcal{E}(S,\mathcal{P},\mu_0)$ *and we shall call them the* points which are $(\mathcal{P},S)$–absolutely continuous with respect to μ_0.

(iii) If Ω *is a compact Riemannian manifold we call the* largest coefficient of expansion of a line element of Ω under the action of S *the quantity*

$$\lambda(S)=\sup_x\frac{\|dSv\|_{V_{Sx}}}{\|v\|_{V_x}},$$

where V_x *is the tangent space to* Ω *at* x *and* $v\in V_x$*.*

Then the following proposition holds.

P3.1.2 **(3.1.2) Proposition:** (Kouchnirenko's theorem)
Let (Ω,S) *be a dynamical system with* Ω *a* C^∞ r*–dimensional compact Riemannian manifold and with* S *a* C^∞ *diffeomorphism of* Ω*. Let* μ_0 *be the volume measure on* Ω *and let* $\mathcal{P}$ *be a* C^∞*–regular partition of* Ω *(see definition (1.4.1)). Let* $\underline{\widehat\sigma}\in\widehat{\mathcal{E}}(S,\mathcal{P},\mu_0)$*.*

Then

e3.1.20
$$s(\underline{\widehat\sigma})\le r\log\lambda, \tag{3.1.20}$$

where $\lambda=\max(\lambda(S),\lambda(S^{-1}))$*, see definition (3.1.2).*

Remarks: (1) The theorem is remarkable because it shows in a general enough context what at first sight might be surprising: namely the entropy of a motion whose initial datum is randomly selected with a probability distribution which is ergodic and equivalent to the volume measure on Ω *cannot exceed* $r\log\lambda$, no matter how fine the partition $\mathcal{P}$ of Ω is taken, provided $\mathcal{P}$ is a regular partition (of course).

(2) Proposition (3.1.2) and certain modifications of it have remarkable applications to the theory of Hamiltonian systems: in studying the latter in connection with Statistical Mechanics one is often interested in motions whose initial data x are randomly chosen, on the energy surface in phase space, with respect to the volume measure.

Consider for instance a Hamiltonian system and select the initial data with the Liouville distribution on an energy surface, or with a distribution equivalent to it. If the system is ergodic (3.1.19) holds (with $\rho=1$) and therefore the complexity of motions observed, say, at unit time intervals and on regular partitions is, with probability 1, bounded by a geometric constant which is independent on the particular motion considered.

Proof: Proceeding as in the proof of proposition (3.1.1) and assuming for the time being, for simplicity, that the sets $P_0, \ldots, P_n$ have diameter small with respect to the diameter of Ω (if not see footnote 6) and of a size such that for all sets of smaller diameter a generalization of the isoperimetric inequality holds, *i.e.*

e3.1.21
$$\mu_0(P) \le \Gamma \,|\, \partial P \,|^{\,r/(r-1)}, \tag{3.1.21}$$

we see that (3.1.14) becomes

e3.1.22
$$\begin{aligned}\sum_{\sigma_0\ldots\sigma_{N-1}} |\, \partial P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}} \,| &\le \sum_{k=0}^{N-1}\sum_{\sigma=0}^{n} |\, S^{-k}\partial P_\sigma \,| \\ &\le 2L\sum_{k=0}^{N-1}\lambda^k = 2L\frac{\lambda^N-1}{\lambda-1}.\end{aligned} \tag{3.1.22}$$

Note that in (3.1.22) one has λ^k rather than $\lambda^{k\,(r-1)}$, as perhaps one might expect, because the conservation of the measure $\rho\,\mu_0$ and the boundedness of ρ and of ρ^{-1} imply that a surface element can at most expand its area by a factor λ.

Hence, by proceeding as in (3.1.17) and (3.1.18), with (3.1.22) instead of (3.1.14), one obtains

e3.1.23
$$\begin{aligned}&\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_2(N)} p\Big(\begin{matrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{matrix}\;\Big|\;\underline{\sigma}\Big) \le \\ &\le e^{-N\eta/r}\lambda^{-N}\Gamma_r^{1-1/r}\,\|\rho\|_\infty^{1-1/r}2L(\lambda^N-1)/(\lambda-1) \le Ge^{-\eta N/r},\end{aligned} \tag{3.1.23}$$

having set

$$\begin{aligned}\mathcal{C}_1(N) &= \Big\{\sigma_0\ldots\sigma_{N-1} \,\Big|\, p\Big(\begin{matrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{matrix}\;\Big|\;\underline{\sigma}\Big) > e^{-\eta N}\lambda^{-Nr}\Big\},\\ \mathcal{C}_2(N) &= \Big\{\sigma_0\ldots\sigma_{N-1} \,\Big|\, p\Big(\begin{matrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{matrix}\;\Big|\underline{\sigma}\Big) \le e^{-\eta N}\lambda^{-Nr}\Big\},\end{aligned}$$

and having chosen G to be a suitable constant.

Since it is clear that the number of elements of $\mathcal{C}_1(N)$ is smaller than
N3.1.6 $e^{N\eta}\lambda^{rN}$ we deduce $s(\widehat{\underline{\sigma}}) \le r\log\lambda$.[6] ■

[6] If the sets have large diameter we can always divide them into smaller sets reducing to the case in which there is at most one set $P^{0,\ldots,N-1}_{0,\ldots,0}$ with a large diameter: and we shall add the latter set to $\mathcal{C}_1$ which will have a number of elements bounded by $e^{N\eta}\lambda^{rN}+1$ and reaching the same conclusions. One makes use of the remark that finer partitions canoot have lower entropy: if $\mathcal{P}'$ is fined than $\mathcal{P}$ then $s(\mathcal{P},S) \le s(\mathcal{P}',S)$.

After introducing the notion of complexity of motions we shall try of analyze the problem of the actual computation of $s(\underline{\sigma}(x))$.

A first remarkable result is the *theorem of Shannon–McMillan* that is fundamental for the major conceptual clarification that it introduces about the meaning of the entropy notion and for its applications to the theory of codes: we shall touch this subject also at the end of Chapter 10 and in some problems in the coming sections.

Problems for §3.1

Q3.1.1 **[3.1.1]:** Construct a sequence in $\{0,1\}^{\mathbb{Z}}$ such that all strings of length $\le N_0$ have frequencies well defined and positive.

Q3.1.2 **[3.1.2]:** Construct a sequence in $\{0,1\}^{\mathbb{Z}}$ with well defined frequencies.

Q3.1.3 **[3.1.3]:** Construct a sequence in $\{0,1\}^{\mathbb{Z}}$ in which all possible strings of finite length appear at least once but have frequency zero except those with specification $\sigma_0 = \sigma_1 = \ldots = 0$.

Q3.1.4 **[3.1.4]:** (*Finite algorithms are simple*) Let $k > 0$ and $f : \{0,\ldots,n\}^k \to \{0,\ldots,n\}$ and define arbitrarily $\sigma_{-1}, \sigma_{-2}, \ldots$; set, inductively for $i \ge 0$, $\sigma_i = f(\sigma_{i-1},\ldots,\sigma_{i-k})$. Show that $s(\underline{\sigma}) = 0$: *i.e.* "is not possible construct complex sequences with finite algorithms".

Q3.1.5 **[3.1.5]:** If $\mathcal{P}$ is an analytically regular partition of $\mathbb{T}^r$, $\underline{\omega} \in \mathbb{R}^r$ and $S\underline{\varphi} = \underline{\varphi} + \underline{\omega} \bmod 2\pi$ show that $s_{\rm abs}(\underline{\varphi}) = 0$ (*Hint:* Consider first the case of the irrational rotations).

Q3.1.6 **[3.1.6]:** (*Typical entropy in a Bernoulli shift*)
A randomly selected sequence in $\{0,1\}^{\mathbb{Z}}$ with an equal weights Bernoulli distribution $B(1/2,1/2)$ has well defined frequencies and entropy $\log 2$: show this by applying (3.1.6) and a combinatorial argument.

Q3.1.7 **[3.1.7]:** Study the problem analogous to problem [3.1.6] for the Bernoulli scheme $B(1/3,2/3)$. The result is $s(\underline{\sigma}) = -(1/3)\log(1/3) - (2/3)\log(2/3))$. (*Hint:* Let $m(\underline{\sigma}) = N^{-1}\sum_{i-0}^{N}\sigma_i$. For any ε one can set $\mathcal{C}_2 = \{\underline{\sigma}|\,|m(\underline{\sigma}) - 2/3| > \delta\}$ chosing δ in a suitable way.)

Q3.1.8 **[3.1.8]:** Generalize problem [3.1.8] to the case of the Bernouilli scheme $B(\pi_0,\ldots,\pi_n)$ on $\Omega = \{0,\ldots,n\}^{\mathbb{Z}}$.

Q3.1.9 **[3.1.9]:** Consider the sequence obtained by writing n_0 zeroes followed by n_0 ones followed by n_1 zeroes followed by n_1 ones, *etc*where $(n_0 + n_1 + \ldots + n_k)^{-1} n_{k+1} \xrightarrow[k\to\infty]{} 0$. Show that the sequence has well defined frequencies and compute the associated distribution $\underline{p}$ and the entropy.

Q3.1.10 **[3.1.10]:** (*A complex sequence generated by a circle rotation*)
Let $\underline{\sigma}$ be a sequence in $\{0,1\}^{\mathbb{Z}}$ with well defined frequencies distributed as the Bernoulli scheme $B(1/2,1/2)$ (see problem [3.1.6]). Let S an irrational rotation of $\mathbb{T}^1$: $S\varphi = \varphi + \omega \bmod 2\pi$. Consider the trajectory of the origin $(S^k 0)_{k\in\mathbb{Z}}$. Associate with the point $S^k 0$ the symbol σ_k and construct two Borel sets $P_0 = \{x\,|\,x = S^k 0 \text{ for } k \text{ with } \sigma_k = 0\}$ and $P_1 = \mathbb{T}^1 \backslash P_0$. Show that the entropy of the motion of 0 observed on the partition $\mathcal{P}$ is $\log 2$ and that this does not contradict proposition (3.1.1). (*Hint:* The partition is not analytically regular).

Q3.1.11 **[3.1.11]:** (*Arbitrarily complex sequence generated by a quasi periodic motion*)
If S is an irrational rotation of the circle $\mathbb{T}^1$, given $\varphi \in \mathbb{T}^1$ and $M > 0$, there exists a Borel partition $\mathcal{P}$ of $\mathbb{T}^1$ on which the motion of φ appears with an entropy larger than M. Why this is not in contradiction with proposition (3.1.2)? (*Hint:* At the light of problem [3.1.10] consider a Bernoulli scheme with $\exp M$ symbols.)

Bibliographical note to §3.1

The notion of complexity goes back to Shannon, [Sh49]. Proposition (3.1.2) is inspired from Kouchnirenko's theorem in the version that one finds in [AA68], p. 46.

§3.2 The Shannon–McMillan theorem

It is the following proposition.

P3.2.1 **(3.2.1) Proposition:** (Shannon–McMillan theorem)
Let $\underline{\widehat\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}}$ be a sequence with defined frequencies and with distribution $\underline{p} = \left(p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\right)$.
(i) The limit

e3.2.1
$$s = \lim_{N\to\infty} -N^{-1} \sum_{\sigma_0\ldots\sigma_{N-1}} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \tag{3.2.1}$$

exists and will be called average entropy *of $\underline{\widehat\sigma}$,*
(ii) If $\underline{\widehat\sigma}$ is ergodic then

e3.2.2
$$s(\underline{\widehat\sigma}) = s \tag{3.2.2}$$

and (3.2.1) often yields a rather convenient way of computing $s(\underline{\widehat\sigma})$.
(iii) If $\underline{\widehat\sigma}$ is ergodic, given $\varepsilon > 0$, there exists N_ε such that for all $N \ge N_\varepsilon$ the elements of $\{0,\ldots,n\}^N$ can be split into two classes $\mathcal{C}_{1,\varepsilon}(N)$ and $\mathcal{C}_{2,\varepsilon}(N)$ with the properties

e3.2.3
$$\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{2,\varepsilon}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} < \varepsilon, \tag{3.2.3}$$

e3.2.4
$$\exp((s-\varepsilon)N) \le |\mathcal{C}_{1,\varepsilon}(N)| \le \exp((s+\varepsilon)N), \tag{3.2.4}$$

e3.2.5
$$\exp(-(s+\varepsilon)N) \le p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le \exp(-(s-\varepsilon)N), \tag{3.2.5}$$

for every choice of the string $\sigma_0\ldots\sigma_{N-1} \in \mathcal{C}_{1,\varepsilon}(N)$. Here $|\mathcal{C}_{1,\varepsilon}(N)|$ denotes the number of elements in $\mathcal{C}_{1,\varepsilon}(N)$.

Remarks: (1) Hence if $\underline{\widehat\sigma}$ is ergodic the strings of symbols of large length can be divided into two groups, one with small total probability and another consisting of "few" elements of approximately equal frequency of appearance in $\underline{\widehat\sigma}$ (in the rather weak sense of (3.2.5)).
(2) For a better understanding of the meaning of the entropy notion one can remark that, if $\underline{\widehat\sigma}$ is ergodic, a quantity s enjoying the properties described in (iii) is necessarily equal to the entropy of $\underline{\widehat\sigma}$.

Indeed if $s(\underline{\widehat\sigma}) = s'$ it is clear that (3.2.3) and (3.2.4) imply that $s' \le s$.

Then suppose that one has $s' < s$. Then there is an ε with $0 < \varepsilon < \min(1/4, (s-s')/4)$ and one can find ε' and a set $\overline{\mathcal{C}}_{1,\varepsilon'}(N)$ in $\{0,\ldots,n\}^N$ such that

e3.2.6
$$|\,\overline{\mathcal{C}}_{1,\varepsilon'}(N)\,| < \exp((s'+\varepsilon)N) \qquad \text{and} \qquad \sum_{\sigma_0\ldots\sigma_{N-1}\notin\overline{\mathcal{C}}_{1,\varepsilon'}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} < \varepsilon', \tag{3.2.6}$$

for infinitely many values of $N > N_{\varepsilon'}$, by the definition of entropy (cf. definition (3.1.1) and equation (3.2.5)), at least if ε is small enough.

The set $\overline{\mathcal{C}}_{1,\varepsilon'}(N)$ will contain ν elements of $\mathcal{C}_{1,\varepsilon}(N)$, $\nu \ge 0$, together with some others of $\mathcal{C}_{2,\varepsilon}(N)$. Hence $\overline{\mathcal{C}}_{1,\varepsilon'}(N)$ will be obtained from $\mathcal{C}_{1,\varepsilon}(N)$ subtracting from it $(|\,\mathcal{C}_{1,\varepsilon}(N)\,| - \nu)$ elements and adding to it a suitable number of other elements of $\mathcal{C}_{2,\varepsilon}(N)$.

But if N is such that (3.2.5) holds for it and if we note the inclusion $\mathcal{C}_{1,\varepsilon}(N)\backslash(\mathcal{C}_{1,\varepsilon}(N)\cap\overline{\mathcal{C}}_{1,\varepsilon'}(N)) \subset \overline{\mathcal{C}}_{2,\varepsilon'}(N)$, we get

e3.2.7
$$\begin{aligned}\varepsilon \ge &\sum_{(\sigma_0\ldots\sigma_{N-1})\in\mathcal{C}_{1,\varepsilon}(N)\backslash(\mathcal{C}_{1,\varepsilon}(N)\cap\overline{\mathcal{C}}_{1,\varepsilon'}(N))} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} = \\ &= \sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} - \\ &- \sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon}(N)\cap\overline{\mathcal{C}}_{1,\varepsilon'}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\end{aligned} \tag{3.2.7}$$

and by (3.2.3), (3.2.5) the last difference can be bounded below by

e3.2.8
$$\begin{aligned}\varepsilon \ge\, &1-\varepsilon-\nu \max_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \ge \\ &\ge 1-\varepsilon-\nu e^{-(s-\varepsilon)N} \ge 1-\varepsilon-e^{(N(s'+\varepsilon)-N(s-\varepsilon))} \ge \\ &\ge 1-\varepsilon-e^{(s'-s+2\varepsilon)N} \ge 1-\varepsilon-e^{-2N\varepsilon},\end{aligned} \tag{3.2.8}$$

being $\nu \le |\,\overline{\mathcal{C}}_{1,\varepsilon}(N)\,| \le e^{(s'+\varepsilon)N}$ and $2\varepsilon < (s-s')/2$, $\varepsilon < 1/4$, at least if N is large enough. Hence the contradiction in (3.2.8) implies that the number s with the property (3.2.3), (3.2.4) and (3.2.5), if it exists, is necessarily the entropy of $\underline{\widehat{\sigma}}$.

(3) It is convenient to break the proof into a few lemmas.

L3.2.1 **(3.2.1) Lemma:** *Let $\underline{p}$ be the distribution of $\underline{\widehat{\sigma}}$ and let $m_{\underline{p}}$ be the probability distribution associated with it. If the function of $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}}$*

e3.2.9
$$\underline{\sigma} \;\to\; f_N(\underline{\sigma}) = -N^{-1}\log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \tag{3.2.9}$$

is such that the limit

e3.2.10
$$\widetilde{s}(\underline{\sigma}) \stackrel{def}{=} \lim_{N\to\infty} f_N(\underline{\sigma}) \quad \text{in } L_1(m_{\underline{p}}) \tag{3.2.10}$$

exists, then proposition (3.2.1) follows.

N3.2.1 *Proof of lemma (3.2.1):* By integrating[1] both sides of (3.2.10) we obtain (3.2.1) so that existence of the limit implies (i) of proposition (3.2.1).

The function $\widetilde{s}(\underline{\sigma})$, if the limit defining it exists almost everywhere, is translation invariant, *i.e.* $\widetilde{s}(\underline{\sigma}) = \widetilde{s}(\tau\underline{\sigma})$ holds $m_{\underline{p}}$–almost everywhere. In fact the monotonicity of the logarithm and the compatibility property fulfilled by the p's imply that $\widetilde{s}(\underline{\sigma}) \le \widetilde{s}(\tau\underline{\sigma})$ since

e3.2.11
$$\begin{aligned}\widetilde{s}(\underline{\sigma}) &= \lim_{N\to\infty} -\frac{1}{N}\log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le\\ &\le \lim_{N\to\infty} -\frac{1}{N}\log p\begin{pmatrix}1\ldots N-1\\ \sigma_1\ldots\sigma_{N-1}\end{pmatrix} = \widetilde{s}(\tau\underline{\sigma}),\end{aligned}\tag{3.2.11}$$

hence, integrating this inequality and using the invariance of $m_{\underline{p}}$ (which implies the opposite inequality $\widetilde{s}(\tau\sigma) \le \widetilde{s}(\sigma)$), one deduces that $m_{\underline{p}}$–almost everywhere one has $\widetilde{s}(\underline{\sigma}) = \widetilde{s}(\tau\underline{\sigma})$.

Then if $\widehat{\underline{\sigma}}$ is ergodic also $m_{\underline{p}}$ is such, and therefore $\widetilde{s}(\underline{\sigma}) = \text{constant} = s$, $m_{\underline{p}}$–almost everywhere. In this case the convergence in (3.2.10), which implies convergence in $m_{\underline{p}}$-measure, implies also that for all $\varepsilon > 0$, there exist N_ε and, for all $N \ge N_\varepsilon$, a set $E_{\varepsilon,N} \subset \{0,\ldots,n\}^{\mathbb{Z}}$ such that

e3.2.12
$$\begin{aligned} m_{\underline{p}}(E_{\varepsilon,N}) &< \varepsilon,\\ |f_N(\underline{\sigma}) - s| &< \varepsilon \qquad \text{for all } \underline{\sigma} \notin E_{\varepsilon,N},\end{aligned}\tag{3.2.12}$$

and it is clear that, since f_N is measurable on the algebra of the cylinders with base $[0, N-1]$, (*i.e.* it only depends on $\sigma_0,\ldots,\sigma_{N-1}$), then also $E_{\varepsilon,N}$ can be chosen measurable on this algebra and, therefore, it is a union of cylinders with base $[0, N-1]$. We shall then set

e3.2.13
$$\begin{aligned}\mathcal{C}_{2,\varepsilon}(N) &= \{\sigma_0\ldots\sigma_{N-1} | C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}} \subset E_{\varepsilon,N}\},\\ \mathcal{C}_{1,\varepsilon}(N) &= \{\sigma_0\ldots\sigma_{N-1} | C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}} \cap E_{\varepsilon,N} = \emptyset\} =\\ &= \{0,\ldots,n\}^N \backslash \mathcal{C}_{2,\varepsilon}(N),\end{aligned}\tag{3.2.13}$$

and, therefore, (3.2.12) implies (3.2.3) and (3.2.5). Then (3.2.4) follows from (3.2.5) and from $\sum_{\sigma_0\ldots\sigma_{N-1}} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} = 1$, and (3.2.2) follows from

[1] *i.e.* my multiplying both sides times $p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$ and summing over $\sigma_0\ldots\sigma_{N-1}$.

(3.2.10) by integrating both sides (indeed the limit (3.2.10) takes place in $L_1(m_{\underline{p}})$ as well), and using the remark (2), above. ∎

Hence proposition (3.2.1) follows from the $L_1(m_{\underline{p}})$–convergence of the limit (3.2.10). To prove convergence we consider, for $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ and $j \geq 1$, the functions

e3.2.14
$$\varphi_j(\underline{\sigma}) = -\log \frac{p\begin{pmatrix} -j \ldots -1 & 0 \\ \sigma_{-j} \ldots \sigma_{-1} & \sigma_0 \end{pmatrix}}{p\begin{pmatrix} -j \ldots -1 \\ \sigma_{-j} \ldots \sigma_{-1} \end{pmatrix}}, \tag{3.2.14}$$

which are non-negative (possibly $+\infty$) and $m_{\underline{p}}$–measurable.

L3.2.2 **(3.2.2) Lemma:** *If the limit*

e3.2.15
$$\varphi(\underline{\sigma}) = \lim_{j\to\infty} \varphi_j(\underline{\sigma}) \qquad \text{in} \quad L_1(m_{\underline{p}}) \tag{3.2.15}$$

exists, then proposition (3.2.1) follows.

Remark: In fact we shall also show that such a limit is also reached $m_{\underline{p}}$–almost everywhere.

Proof of lemma (3.2.2): Consider the following identity, for $N \geq 2$,

e3.2.16
$$\begin{aligned} &-N^{-1}\log p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} = \\ &= -N^{-1}\sum_{j=0}^{N-2} \log \frac{p\begin{pmatrix} 0 \ldots j+1 \\ \sigma_0 \ldots \sigma_{j+1} \end{pmatrix}}{p\begin{pmatrix} 0 \ldots j \\ \sigma_0 \ldots \sigma_j \end{pmatrix}} - N^{-1}\log p\begin{pmatrix} 0 \\ \sigma_0 \end{pmatrix}, \end{aligned} \tag{3.2.16}$$

and note that $-N^{-1}\log p\begin{pmatrix} 0 \\ \sigma_0 \end{pmatrix} \xrightarrow[N\to\infty]{} 0$ both in $L_1(m_{\underline{p}})$ and $m_{\underline{p}}$–almost
N3.2.2 everywhere[2].

Furthermore the sum in (3.2.16) can be written, when $N \to \infty$,

e3.2.17
$$N^{-1}\sum_{j=1}^{N-1} \varphi_j(\tau^j \underline{\sigma}), \tag{3.2.17}$$

because, by the translation invariance of $\underline{p}$,

e3.2.18
$$\varphi_j(\tau^j\underline{\sigma}) = -\log \frac{p\begin{pmatrix} -j \ldots -1 & 0 \\ \sigma_0 \ldots \sigma_{j-1} & \sigma_j \end{pmatrix}}{p\begin{pmatrix} -j \ldots -1 \\ \sigma_0 \ldots \sigma_{j-1} \end{pmatrix}} = -\log \frac{p\begin{pmatrix} 0 \ldots j \\ \sigma_0 \ldots \sigma_j \end{pmatrix}}{p\begin{pmatrix} 0 \ldots j-1 \\ \sigma_0 \ldots \sigma_{j-1} \end{pmatrix}}. \tag{3.2.18}$$

---

[2] Remark that if $p\begin{pmatrix} 0 \\ \bar\sigma_0 \end{pmatrix} = 0$ for some $\bar\sigma_0$ the statement is still valid because the $\underline{\sigma}$ with $\sigma_0 = \bar\sigma_0$ have 0 measure.

Existence of the limit (3.2.15) implies, by the τ–invariance of $m_{\underline{p}}$,

e3.2.19 $$N^{-1}\sum_{j=1}^{N-1}(\varphi_j(\tau^j\underline{\sigma})-\varphi(\tau^j\underline{\sigma}))\xrightarrow[N\to\infty]{}0\quad\text{in}\quad L_1(m_{\underline{p}}). \tag{3.2.19}$$

Hence, by Birkhoff's theorem, the limit as $N\to\infty$ exists in $L_1(m_{\underline{p}})$:

e3.2.20 $$\overline{\varphi}(\underline{\sigma})=\lim_{N\to\infty}N^{-1}\sum_{j=1}^{N-1}\varphi(\tau^j\underline{\sigma}) \tag{3.2.20}$$

and, therefore, it also follows $f_N(\underline{\sigma})\xrightarrow[N\to\infty]{}\overline{\varphi}(\underline{\sigma})$ in $L_1(m_{\underline{p}})$, and the function $\widetilde{s}(\underline{\sigma})$ will be just $\overline{\varphi}(\underline{\sigma})$. ∎

We can proceed to conclude the proof of the proposition.

Proof of proposition (3.2.1): The functions φ_j are a set equibounded in
N3.2.3 $L_1(m_{\underline{p}})$ and equisummable[3] with respect to $m_{\underline{p}}$. Indeed let $E_{j,k}$ be the set of the sequences $\underline{\sigma}\in\{0,\ldots,n\}^{\mathbb{Z}}$ such that

e3.2.21 $$k\le\varphi_j(\underline{\sigma})<k+1. \tag{3.2.21}$$

For the sequences $\underline{\sigma}$ contained in such a set one has

e3.2.22 $$p\begin{pmatrix}-j\ldots0\\ \sigma_{-j}\ldots\sigma_0\end{pmatrix}\le e^{-k}\,p\begin{pmatrix}-j\ldots-1\\ \sigma_{-j}\ldots\sigma_{-1}\end{pmatrix}. \tag{3.2.22}$$

Adding up these relations, summing over the choices of $\sigma_{-j},\ldots,\sigma_0\in\{0,\ldots,n\}^{j+1}$ such that $C^{-j\ldots0}_{\sigma_{-j}\ldots\sigma_0}\subset E_{j,k}$, one finds

e3.2.23 $$m_{\underline{p}}(E_{j,k})\le(n+1)e^{-k}, \tag{3.2.23}$$

hence, if E is a $m_{\underline{p}}$–measurable set,

$$\int_E\varphi_j(\underline{\sigma})m_{\underline{p}}(d\underline{\sigma})=\sum_{k=0}^{\infty}\int\varphi_j(\underline{\sigma})\chi_E(\underline{\sigma})\chi_{E_{j,k}}(\underline{\sigma})m_{\underline{p}}(d\underline{\sigma})\le$$

$$\le\sum_{k=0}^{\infty}(1+k)\int\chi_E(\underline{\sigma})\chi_{E_{j,k}}(\underline{\sigma})m_{\underline{p}}(d\underline{\sigma})\le\sqrt{m_{\underline{p}}(E)}\sum_{k=0}^{\infty}(1+k)\sqrt{m_{\underline{p}}(E_{j,k})}\le$$

e3.2.24 $$\le\sqrt{m_{\underline{p}}(E)}\Big[\sqrt{(n+1)}\sum_{k=0}^{\infty}(1+k)e^{-k/2}\Big], \tag{3.2.24}$$

[3] A set of functions $F\subset L_1(\mu)$ is called *equisummable* if for every ε there is a δ such that if $\mu(E)\le\delta$ then $\int_E f(x)d\mu(x)\le\varepsilon$ for every $f\in F$.

that shows simultaneously (and "miracolously") the equiboundedness in $L_1(m_{\underline{p}})$ and the equisummability of φ_j, $j = 1, \ldots$.

Simple considerations of measure theory based on Vitali's convergence theorem and on Fatou's lemma[4] show that, to verify the convergence, in $L_1(m_{\underline{p}})$
N3.2.4 and almost everywhere, of the sequence $\{\varphi_j\}$, as $j \to \infty$, it suffices to verify convergence in $m_{\underline{p}}$–measure and, respectively, almost everywhere of the sequence of functions

e3.2.25
$$\exp(-\varphi_j(\underline{\sigma})) = \frac{p(C^{-j\ldots 0}_{-j\ldots\sigma_0})}{p(C^{-j\ldots -1}_{-j\ldots\sigma_{-1}})} \quad \text{for} \quad j \to \infty, \tag{3.2.25}$$

N3.2.5 as one can check.[5]

The problem of convergence in $m_{\underline{p}}$–measure and $m_{\underline{p}}$–almost everywhere, of $\exp(-\varphi_j)$ is very similar to the problem of proving the *Vitali–Lebesgue theorem* which states the existence of the limit as $I \to y$ of $|I|^{-1}\int_I F(x)dx$ if I are intervals containing y. In the present context the result is called *Doob's theorem*.

This result is a particular case of a general theorem of measure theory, and it is worth to make a small notational effort to reduce it to this general theorem.

All the functions $\exp(-\varphi_j)$, $j = 1, 2, \ldots$ are measurable with respect to the σ–algebra $\mathcal{B}$–generated by the cylinders with negative base; it will be useful
N3.2.6 to identify, in the obvious way, such cylinders with those of [6] $\{0, \ldots, n\}^{\mathbb{Z}_-}$ so that

e3.2.26
$$m(C^J_{\underline{\sigma}}) = m_{\underline{p}}(C^J_{\underline{\sigma}}) = p\begin{pmatrix} J \\ \underline{\sigma} \end{pmatrix} \qquad \forall\, J \subset \mathbb{Z}_-,\ \forall\, \underline{\sigma} \in \{0, \ldots, n\}^{|J|}; \tag{3.2.26}$$

in other words m coincides with $m_{\underline{p}}$ restricted to the cylinders witn negative
N3.2.7 base. By the Radon–Nykodim theorem,[7] given $\sigma_0 \in \{0, \ldots, n\}$, there exists a function in $L_1(m)$, that we shall denote g_{σ_0}, such that

e3.2.27
$$p\begin{pmatrix} -j \ldots -1 0 \\ \sigma_{-j} \ldots \sigma_{-1}\sigma_0 \end{pmatrix} = \int_{C^{-j\ldots -1}_{\sigma_{-j}\ldots\sigma_{-1}}} g_{\sigma_0}(\underline{\sigma}')\, m(d\underline{\sigma}'). \tag{3.2.27}$$

---

[4] See [DS58], p.150 and 152.

[5] If $\exp(-\varphi_j)$ converges in $m_{\underline{p}}$–measure and almost everywhere to a limit that we denote $\exp(-\varphi)$ one has $0 \le \varphi_j \xrightarrow[j\to\infty]{} \varphi$ almost everywhere. By Fatou's lemma φ is summable and hence $< +\infty$ almost everywhere; then convergence in $m_{\underline{p}}$–measure of $\exp(-\varphi_j)$ to $\exp(-\varphi)$ implies convergence in measure of φ_j to φ and therefore, given the equisummability of the functions φ_j, Vitali's criterion of convergence implies the convergence in $L_1(m_{\underline{p}})$ of φ_j to φ.

[6] $\mathbb{Z}_+$ denotes the integers ≥ 0 and $\mathbb{Z}_-$ denotes the integers < 0. Hence $\{0, \ldots, n\}^{\mathbb{Z}_-}$ are unilateral sequences with negative labels.

[7] See [DS58], p.176.

Hence

e3.2.28
$$\frac{p\begin{pmatrix} -j\dots-1 & 0\\ \sigma_{-j}\dots\sigma_{-1} & \sigma_0\end{pmatrix}}{p\begin{pmatrix} -j\dots-1\\ \sigma_{-j}\dots\sigma_{-1}\end{pmatrix}}=\frac{\int_{C^{-1\dots-1}_{\sigma_{-j}\dots\sigma_{-1}}} g_{\sigma_0}(\underline{\sigma}')\,m(d\underline{\sigma}')}{\int_{C^{-j\dots-1}_{\sigma_{-j}\dots\sigma_{-1}}} m(d\underline{\sigma}')},\tag{3.2.28}$$

and we realize that our purpose is to show the convergence, in m–measure and m–almost everywhere, of this quantity regarded as a function of $\underline{\sigma}\in\{0,\dots,n\}^{\mathbb{Z}_-}$, for every fixed σ_0. And the limit should be precisely g_{σ_0}.

It is now possible to reduce the analysis to some simple and classical constructions.

For simplicity we shall suppose that the measures $m_{\underline{p}}$ and m are not atomic.

Let us consider the map of $\{0,\dots,n\}^{\mathbb{Z}_-}$ in $[0,n/n{+}1]$ defined by $\underline{\sigma}\to X(\underline{\sigma})$:

e3.2.29
$$X(\underline{\sigma})=\sum_{j=1}^{\infty}\frac{\sigma_{-j}}{(2+n)^j},\tag{3.2.29}$$

which is a homeomorphism between the sequences in $\{0,\dots,n\}^{\mathbb{Z}_-}$ and the subset $X(\{0,\dots,n\}^{\mathbb{Z}_-})$ of $[0,1]$, which is the Cantor set consisting of the numbers of $[0,1]$ whose development in base $(n+2)$ never contains the digit $n+1$.

Then via the map X we can transform the measure m into a measure $\overline{m}$ on $[0,n/(n+1)]$ and the function g_{σ_0} into $\widetilde{g}\in L_1(\overline{m})$ by setting

e3.2.30
$$\begin{aligned}\overline{m}(E)&=m(X^{-1}E),\qquad E\subset\mathcal{B}([0,n/n+1]),\\ \widetilde{g}(x)&=g_{\sigma_0}(X^{-1}(x)).\end{aligned}\tag{3.2.30}$$

Define then the map $Y:[0,n/n+1]\to[0,1]$ as

e3.2.31
$$y=Y(x)=\overline{m}([0,x]).\tag{3.2.31}$$

The function $x\to Y(x)$ is non-decreasing, more precisely it is strictly increasing except, possibly, in the union of a denumerable family of closed disjoint intervals. Hence Y is continuous and is invertible as a map between $[0,n/n+1]$ deprived of a denumerable infinity of closed disjoint sets and its image in $[0,1]$ (which consists in the same $[0,1]$ deprived, at most, of a denumerable infinity of points).

Therefore Y establishes an isomorphism mod 0 between $([0,n/n+1]),\overline{m})$ and $([0,1],\mu)$ where μ is the Lebesgue measure (because of the relation (3.2.31)). Via this isomorphism $\widetilde{g}$ becomes a function $\overline{g}\in L_1(\mu)$.

Another remarkable fact is that the set

e3.2.32
$$D(\sigma_{-j}\dots\sigma_{-1})=YX(C^{-j\dots-1}_{\sigma_{-j}\dots\sigma_{-1}})\tag{3.2.32}$$

either is a connected interval or is empty; the latter possibility arises if and only if $m(C^{-j\dots-1}_{\sigma_{-j}\dots\sigma_{-1}})=0$. Indeed it suffices to remark that the set of the numbers that have the first j digits of the their development in base $(n+1)$ equal is an interval.

The nonempty intervals having the form (3.2.32) form a covering $\mathcal{D}_j$ of $[0,1]$ with intervals which, pairwise, have no internal points in common. Furthermore the covering $\mathcal{D}_{j+1}$ refines $\mathcal{D}_j$ because every interval of $\mathcal{D}_{j+1}$ is a union of intervals of $\mathcal{D}_j$.

If $x \in D(\sigma_{-j},\ldots,\sigma_{-1})$ and if we define the function

e3.2.33
$$x \to h_j(x) = \frac{\int_{D(\sigma_{-j}\ldots\sigma_{-1})} \overline{g}(x')dx'}{|D(\sigma_{-j}\ldots\sigma_{-1})|} \tag{3.2.33}$$

(which is a relation that has μ–almost everywhere meaning) we obtain the image of φ_j via YX.

The Vitali–Lebesgue theorem concerns exactly sequences of functions having the form

e3.2.34
$$x \to q_j(x) = \frac{\int_D g(x')dx'}{|D|} \quad \text{with} \quad g \in L_1(\mu), \tag{3.2.34}$$

where D is the interval that contains x extracted out of a pavement $\mathcal{D}_j$ of $[0,1]$ with intervals with no common internal points and refined by $\mathcal{D}_{j+1}$. The theorem says that $q_j \xrightarrow[j\to\infty]{} g$ almost everywhere with respect to the
N3.2.8 Lebesgue measure μ in $[0,1]$ and in $L_1(\mu)$.[8]

Applying the latter statement to (3.2.33) and translating it back to the original variables via the isomorphism YX we see that it means

e3.2.35
$$\lim_{j\to\infty} \varphi_j(\underline{\sigma}) = g_{\sigma_0}(\underline{\sigma}) \tag{3.2.35}$$

$m_{\underline{p}}$–almost everywhere and in $L_1(m_{\underline{p}})$. This yields the proof of Doob's theorem in the present special case and completes the proof of proposition (3.2.1).

Problems for §3.2

Q3.2.1 **[3.2.1]:** (*Approximability in entropy and distribution*)

Under the hypothesis of proposition (3.2.1) assume $\widehat{\underline{\sigma}}$ ergodic and set $S = s(\widehat{\sigma})$. Show that given an integer $u > 0$ and $\varepsilon > 0$, an integer $N(\varepsilon, u)$ exists such that for all $N > N(\varepsilon, u)$ it is possible to divide $\{0,\ldots,n\}^N$ into two classes $\mathcal{C}_{1,\varepsilon,u}(N)$ and $\mathcal{C}_{2,\varepsilon,u}(N)$ such that

$$\begin{aligned}
&\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{2,\varepsilon,u}(N)} p\begin{pmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le \varepsilon,\\
&\exp((S-\varepsilon)N) \le |\mathcal{C}_{1,\varepsilon,u}(N)| \le \exp[(S+\varepsilon)N],\\
&\exp[-(S+\varepsilon)N] \le p\begin{pmatrix} 0\ldots N-1 \\ \sigma'_0\ldots\sigma'_{N-1}\end{pmatrix} \le \exp[-(S-\varepsilon)N],
\end{aligned}$$

[8] See, for instance, [DS58] Vol. I, Ch. III, p. 214. One should note the "analogy" between the proof discussed here and that of Birkhoff's theorem in Appendix 2.2.

for all $(\sigma'_0 \ldots \sigma'_{N-1}) \in \mathcal{C}_{1,\varepsilon,u}(N)$ and

$$\sum_{\sigma_0 \ldots \sigma_{u-1} \in \{0,\ldots,n\}^u} \Big| p \begin{pmatrix} 0 \ldots u-1 \\ \sigma_0 \ldots \sigma_{u-1} \end{pmatrix} - \text{(frequency of appearance of } \begin{pmatrix} 0 \ldots u-1 \\ \sigma_0 \ldots \sigma_{u-1} \end{pmatrix} \text{ in } (\sigma'_0, \ldots, \sigma'_{N-1}) \Big| < \varepsilon$$

for all $(\sigma'_0 \ldots \sigma'_{N-1}) \in \mathcal{C}_{1,\varepsilon,u}(N)$. (*Hint:* Proceed as in the above derivation of (3.2.3), (3.2.4), (3.2.5) from (3.2.12) observing that, by Birkhoff's theorem,

$$g_N(\underline{\sigma}) = \sum_{\substack{\sigma'_0 \ldots \sigma'_{u-1} \\ \in \{0,\ldots,n\}^u}} p \begin{pmatrix} 0 \ldots u-1 \\ \sigma'_0 \ldots \sigma'_{u-1} \end{pmatrix} - \text{(frequency of appearance of } \begin{pmatrix} 0 \ldots u-1 \\ \sigma'_0 \ldots \sigma'_{u-1} \end{pmatrix} \text{ between 0 and } N-1 \text{ in } \underline{\sigma})$$

converges, for $N \to \infty$, to zero in $L_1(m_{\underline{p}})$ and $m_{\underline{p}}$–almost everywhere. Then choose $E_{\varepsilon,u,N}$ in analogy with the choice (3.2.12) but such that $|g_N(\underline{\sigma})| < \varepsilon$, *etc.*)

Q3.2.2 **[3.2.2]:** (*Entropy of a distribution on symbolic sequences*)
If $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ we can define the entropy of $\underline{p}$ naturally as

$$s(\underline{p}) = \lim_{\varepsilon \to 0} \limsup_{N \to \infty} N^{-1} \log \mathcal{N}_\varepsilon(\underline{p}|N),$$

where $\mathcal{N}_\varepsilon(\underline{p}\,|\,N)$ is the minimum number of elements of $\{0,\ldots,n\}^N$ that remain if we take out from $\{0,\ldots,n\}^N$ a family $\mathcal{C}_2(N)$ of elements such that $\sum_{\sigma_0 \ldots \sigma_{N-1} \in \mathcal{C}_2(N)} p \begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \leq \varepsilon$.

Then proposition (3.2.1) formulated in terms of $\underline{p}$ has a meaning (in an obvious way, even if there is no $\widehat{\underline{\sigma}} \in \{0,\ldots,n\}^{\mathbb{Z}}$ that generates $\underline{p}$) and is true (note that this statement does not strengthen proposition (3.2.1) except for what concerns its statement (i), cf. proposition (3.2.1), and the relative remarks).

Q3.2.3 **[3.2.3]:** (*Lebesgue measure on* $[0,1]^2$ *and the* $(\frac{1}{2},\frac{1}{2})$ *Bernoulli shift*)
Show that the Bernoulli distribution $B(1/2,1/2)$ on $\{0,1\}^{\mathbb{Z}}$ is isomorphic mod 0 to the Lebesgue measure on the square $[0,1]\times[0,1]$. The isomorphism is established by $(x,y) \in [0,1]^2 \to \underline{\sigma} \in \{0,1\}^{\mathbb{Z}}$ if $x = \sum_{j=0}^\infty \frac{\sigma_j}{2^{j+1}}$, $y = \sum_{j=1}^\infty \frac{\sigma_{-j}}{2^j}$.

Q3.2.4 **[3.2.4]:** (*The baker map and the* $(\frac{1}{2},\frac{1}{2})$ *Bernoulli shift*)
The isomorphism of problem [3.2.3] establishes an isomorphism mod 0 between the dynamical systems $(\{0,1\}^{\mathbb{Z}},\tau,B)$ and $([0,1]^2,S,\lambda)$ where $\lambda(dx) = dxdy$ and $S(x,y) = (2x, y/2)$ if $x < 1/2$, and $S(x,y) = (2x-1,(y+1)/2)$ if $x \geq 1/2$. The latter dynamical system is called *the baker map* (see also problem [2.2.43]).

Q3.2.5 **[3.2.5]:** (*Generalization of the binary and decimal expansions*)
Consider n positive numbers $p_1,\ldots,p_n$ such that $\sum_{i=1}^n p_i = 1$. Consider n intervals $I_1, I_2,\ldots,I_n$ that decompose $[0,1)$. Define $Sx = (x-a_i)/p_i$, if $x \in I_i = [a_i,a_{i+1})$, having set $a_0 = 0$ and $a_{i+1} = p_1 + \ldots + p_i$, $i = 1,\ldots,n$. Draw the map S as a map of $[0,1]$ into itself and show that S conserves the Lebesgue measure on $[0,1)$. The code that associates with $x \in [0,1]$ its history on $(I_1,\ldots,I_n) : x \to (\sigma_0,\sigma_1,\ldots,) \in \{1,\ldots,n\}^{\mathbb{Z}_+}$ transforms the Lebesgue measure into the unilateral Bernoulli measure $B(p_1,\ldots,p_n)$ on $\{0,\ldots,n\}^{\mathbb{Z}_+}$. This code generalizes the binary representation (which corresponds to the case $n = 2$ and $p_1 = p_2 = 1/2$).

Q3.2.6 **[3.2.6]:** (*A generalization of the baker map isomorphism with a Bernoulli shift*) Generalize the result of problem [3.2.5] to show that the Bernoulli scheme with n symbols and probabilities $(p_1, \ldots, p_n)$ is isomorphic mod 0 to the Lebesgue measure on $[0,1]^2$ on which acts a suitable map S.

Q3.2.7 **[3.2.7]:** Consider the Bernoulli scheme $B(p_1, \ldots, p_n)$ on $\{1, \ldots, n\}^{\mathbb{Z}}$. Given k positive numbers $a_1, \ldots, a_k$ such that $a_1 + \ldots + a_k = 1$, show the existence of a Borel partition of $\{1, \ldots, n\}^{\mathbb{Z}}$ into k sets of measures, respectively, $a_1,\ a_2, \ldots, a_k$. (*Hint:* Use the result in problem [3.2.5] and the fact that such partitions trivially exist on $[0,1]^2$ considered with the Lebesgue measure).

Q3.2.8 **[3.2.8]:** Given a Borel partition $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ of $\{1, \ldots, n\}^{\mathbb{Z}}$, and given the Bernoulli measure μ on $\{1, \ldots, n\}^{\mathbb{Z}}$ with probabilities $B(p_1, \ldots, p_n)$ show, by making use of the results of problems [3.2.6] and [3.2.7], that there exists a family of partitions $t \to \mathcal{Q}(t)$ parameterized bt $t \in [0,1]$, such that $\mathcal{Q}(0) = \{\emptyset, \emptyset, \ldots, \emptyset, \{0, \ldots, n\}^{\mathbb{Z}}\}$, $\mathcal{Q}(1) = \{Q_1, \ldots, Q_k\}$ with a fixed number, k, of atoms and which is continuous in the sense that

$$\lim_{t \to t_0} |\mathcal{Q}(t), \mathcal{Q}(t_0)| \equiv \lim_{t \to t_0} \sum_{i=1}^{k} \mu(Q_i(t) \Delta Q_i(t_0)) = 0,\ \forall\, t_0 \in [0,1]\ ,$$

where $A \Delta B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference between A and B. (*Hint:* Consider the isomorphism discussed in problem [3.2.6] and use that such a property is easy to show in the case of Borel partitions of $[0,1]^2$).

Q3.2.9 **[3.2.9]:** (*Non-atomic Borel measures on* $\{0,1\}^{\mathbb{Z}}$ *are isomorphic* mod 0 *to the Lebesgue measure on* $[0,1]$)
Let μ be a non-atomic Borel measure on $\{0,1\}^{\mathbb{Z}}$ (*i.e.* a measure such that no positive measure set E exists which has no subsets of smaller but positive measure). Show that it is isomorphic mod 0 to the Lebesgue measure on $[0,1]$. (*Hint:* Use the idea and the map YX that appear at the end of the proof of proposition (3.2.1).)

Q3.2.10 **[3.2.10]:** Show that if μ is an S–invariant measure on a σ–algebra $\mathcal{B}$ of Ω and it is S–mixing then μ is non-atomic if $\mathcal{B}$ is not trivial.

Q3.2.11 **[3.2.11]:** If μ is a non-atomic Borel measure on a complete and separable metric space then μ is isomorphic mod 0 to a Borel measure on $\{0,1\}^{\mathbb{Z}}$. (*Hint:* Use Alexandrov and
N3.2.9 Urhyson theorems[9] stating that each separable and complete metric space is homeomorphic to a Borel subset of $[0,1]^{\mathbb{Z}}$ which, in turn, is in a one-to-one and bimeasurable correspondence with a set Borel of $\{0,1\}^{\mathbb{Z}}$. Hence by using the conclusions of problem [3.2.9] show that μ is isomorphic mod 0 to the Lebesgue measure on $[0,1]$.)

Q3.2.12 **[3.2.12]:** Show that every dynamical system (Ω, S, μ) with Ω complete metric separable and with μ non-atomic is isomorphic mod 0 to a dynamical system having the form $([0,1], \widetilde{S}, \mu_0)$ where μ_0 is the Lebesgue measure and $\widetilde{S}$ is a suitable map.

Q3.2.13 **[3.2.13]:** Interpret proposition (2.3.3) as a proof that every invertible topological dynamical system (Ω, S, μ) on a complete metric space admitting a topological separating partition is isomorphic mod 0 to a system of the type $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, m)$, where τ is the translation of the sequences of symbols.

Q3.2.14 **[3.2.14]:** If (Ω, S, μ) is an invertible dynamical system on Ω that is assumed be a complete metric separable space and if μ is a complete Borel measure (cf. Appendix 1.4), then (Ω, S, μ) is isomorphic mod 0 to a dynamical system of the type $(V^{\mathbb{Z}}, \tau, \widetilde{\mu})$ where V is a compact metric space, τ is the translation on $V^{\mathbb{Z}}$, and $\widetilde{\mu}$ is a τ–invariant complete Borel measure. (*Hint:* Let $x_1,\ x_2, \ldots$ be a denumerable dense set and consider the function $\varphi : \Omega \to [0,1]^{\mathbb{Z}_+}$ defined by $\varphi(x) = (d(x, x_i)/1 + d(x, x_i))_{i \in \mathbb{Z}}$, if $d(\ldots)$ is the metric on Ω. Then φ is an isomorphism between Ω and its image $\varphi(\Omega) \subset [0,1]^{\mathbb{Z}_+}$ that turns out

[9] See [DS58] pp. 24,138, and problem [3.2.14].

to be a Borel set, and in fact a G_δ–set (*i.e.* a countable intersection of open dense sets),
N3.2.10 in the product topology (theorem, of Alexandrov and Urhyson[10]). Associate then with x the sequence in $V^{\mathbb{Z}} = ([0,1]^{\mathbb{Z}_+})^{\mathbb{Z}}$ defined by $\Phi(x) = (\varphi(S^i x))_{i\in\mathbb{Z}}$; it is clear that Φ is a continuous and one-to-one map between Ω and $\varphi(\Omega)$; hence it is bimeasurable (by
N3.2.11 Kuratowsky's theorem[11] This implies that if we set $\mu(E) = \mu(\Phi^{-1}(E))$ for E Borel in $V^{\mathbb{Z}}$, then $(V^{\mathbb{Z}}, \tau, \widetilde{\mu})$ is isomorphic mod0 to (Ω, S, μ).)

One says, therefore, that "every metric invertible dynamical system constructed on a complete separable metric space by means of a map and of a Borel measure is isomorphic mod (0) to a topological dynamical system on a compact metric space".

Bibliographical note to §3.2

The proof of proposition (3.2.1) ("Shannon–McMillan theorem") is taken from [Ki57], p. 44–89.

The relation between Doob's theorem and Vitali–Lebesgue's is well known. The proof of Doob's theorem can be found in [Ki57]. A proof of the Vitali–Lebesgue theorem can be found, for instance, in [DS58], p.214.

We remark that for a proof of proposition (3.2.1) the L_1–convergence of (3.2.15) would be sufficient. Instead, we have also obtained (implicitly) the almost everywhere convergence: this provides a strenghtening of the Shannon–McMillan theorem (due to Breiman, cf. [Br57]).

A generalization of the notion of entropy to measures on $\{0,\ldots,n\}^{\mathbb{Z}}$ that are not invariant under translation can be found in [Ja59], where an extension of the theorem of Shannon–McMillan to "S-quasi–periodic" distributions (rather than S-invariant) is discussed.

§3.3 Elementary properties of the average entropy

In this section some consequences of the proof of proposition (3.2.1) are collected together with some elementary properties of entropy and with various interesting definitions.

C3.3.1 **(3.3.1) Corollary:** (Average entropy of a sequence)
If $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ is a stationary distribution on $\{0,\ldots,n\}^{\mathbb{Z}}$ the function defined by the limit

e3.3.1
$$\widetilde{s}(\underline{\sigma}) = \lim_{N\to\infty} -N^{-1} \log p \begin{pmatrix} O\ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \tag{3.3.1}$$

exists in $L_1(m_{\underline{p}})$ and (therefore) the limit

e3.3.2
$$\widetilde{s}(\underline{p}) = \lim_{N\to\infty} -N^{-1} \sum_{\sigma_0\ldots\sigma_{N-1}} p \begin{pmatrix} 0\ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \log p \begin{pmatrix} 0\ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \tag{3.3.2}$$

---

[10] It is a useful exercise to look for a proof of this theorem without having recourse to the literature.

[11] See footnote 3, Section §2.3.

exists.

Proof: In the proof of statement (i) of proposition (3.2.1) (*i.e.* in lemma (3.2.1) needed to prove it) the sequence $\underline{\widehat{\sigma}}$ generating the distribution of frequencies $\underline{p}$ enters only through $m_{\underline{p}}$. And we only made use of the translation invariance of $m_{\underline{p}}$: the latter property is also true for the distribution $\underline{p}$ that appears in (3.3.1) even if $\underline{p}$ is not the distribution of a sequence. ■

D3.3.1 **(3.3.1) Definition:** *If $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ the quantity $\widetilde{s}(\underline{p})$ in (3.3.2) will be called* average entropy *of $\underline{p}$. The notion is a priori different from that of* entropy *of $\underline{p}$, that we shall denote by $s(\underline{p})$, which can be defined by generalizing definition (3.1.1) to the case in which $\underline{p}$ is not generated by a*
N3.3.1 *sequence $\underline{\sigma}$.*[1]

Remarks: (1) If $\underline{p}$ is ergodic $m_{\underline{p}}$ is also ergodic and, therefore, $\widetilde{s}(\underline{\sigma})$ is almost everywhere constant (with respect to $m_{\underline{p}}$, because we have seen in the proof of lemma (3.2.1) that it is translation invariant) and one can repeat the first part of the proof of proposition (3.2.1) to conclude that for all $\varepsilon > 0$, there is N_ε such that for $N \geq N_\varepsilon$ the elements of $\{0,\ldots,1\}^N$ can be divided in two sets $\mathcal{C}_{1,\varepsilon}(N)$ and $\mathcal{C}_{2,\varepsilon}(N)$ satisfying (3.2.3), (3.2.4) and (3.2.5).

We can then repeat the arguments of remark (1) to proposition (3.2.1) and deduce that $\widetilde{s}(\underline{\sigma}) = s = s(\underline{p})$. Hence if $\underline{p}$ is ergodic we have

e3.3.3
$$\widetilde{s}(\underline{p}) = s(\underline{p}). \tag{3.3.3}$$

(2) The relation (3.3.3) does not hold in general if $\underline{p}$ is not ergodic.

The following proposition clarifies the remark (2) above.

P3.3.1 **(3.3.1) Proposition:** (Entropy and average entropy)
If $\underline{p}_1, \underline{p}_2 \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ and are ergodic and, if $\underline{p} = a\underline{p}_1 + (1-a)\underline{p}_2$ with $0 < a < 1$, in the sense that

e3.3.4
$$p\begin{pmatrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix} = a p_1 \begin{pmatrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix} + (1-a) p_2 \begin{pmatrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix}, \tag{3.3.4}$$

one has

e3.3.5
$$\begin{aligned} \widetilde{s}(\underline{p}) &= a s(\underline{p}_1) + (1-a) s(\underline{p}_2), \\ s(\underline{p}) &= \max\{s(\underline{p}_1), s(\underline{p}_2)\}. \end{aligned} \tag{3.3.5}$$

The first property is called affinity *of the entropy.*

---

[1] In other words we can define the *complexity* with weight $e^{-\underline{V}}$ of a shift invariant distribution $\underline{p}$ the quantity $s(\underline{p},\underline{V})$ as in (3.1.8), (3.1.9) and (3.1.10), where $\mathcal{C}_1$ and $\mathcal{C}_2$ are now defined by replacing the frequencies $p\left(\begin{smallmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{smallmatrix} \,|\, \widehat{\sigma}\right)$ by $p\left(\begin{smallmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{smallmatrix}\right)$ and by replacing $\eta_\varepsilon(\underline{\widehat{\sigma}}|N;\underline{V})$ with $\eta_\varepsilon(\underline{p}|N;\underline{V})$, where one defines $\eta_\varepsilon(\underline{p}|N;\underline{V})$ through the second member of (3.1.9). Then $s(\underline{p})$ is naturally defined as $s(\underline{p}\,|\,\underline{0})$, using the notations of definition (3.1.1) with $V_N \equiv 0$.

Proof: The first of (3.3.5) is based on the identity (3.3.3) for ergodic distributions and on the following simple inequalities.

The function $\xi \to -\xi \log \xi$ is convex for $\xi \in [0,1]$ and, therefore,

e3.3.6
$$-(ax+(1-a)y)\log(ax+(1-a)y) \ge -ax\log x-(1-a)y\log y \tag{3.3.6}$$

for every a, x, $y \in [0,1]$. Furthermore if $0 \le x_1,\ldots,x_p \le 1$, $0 \le y_1,\ldots,y_p \le 1$ and $\sum_i x_i = \sum_i y_i = 1$ the monotonicity of $\xi \to -\log\xi$ gives

e3.3.7
$$\begin{aligned}
\sum_i &-(ax_i+(1-a)y_i)\log(ax_i+(1-a)y_i) \equiv\\
&\equiv \sum_i -ax_i\log(ax_i+(1-a)y_i)+\\
&\quad+\sum_i -(1-a)y_i\log(ax_i+(1-a)y_i) \le\\
&\le \sum_i -ax_i\log ax_i + \sum_i -(1-a)y_i\log(1-a)y_i =\\
&= -a\log a-(1-a)\log(1-a)+\\
&\quad + a\sum_i -x_i\log x_i + (1-a)\sum_i -y_i\log y_i .
\end{aligned}\tag{3.3.7}$$

Hence (3.3.6) and (3.3.7) imply

e3.3.8
$$\begin{aligned}
a&\sum_i -x_i\log x_i + (1-a)\sum_i y_i\log y_i \le\\
&\le \sum_i -(ax_i+(1-a)y_i)\log(ax_i+(1-a)y_i) \le\\
&\le -a\log a-(1-a)\log(1-a)+a\sum_i -x_i\log x_i+(1-a)\sum_i -y_i\log y_i ,
\end{aligned}\tag{3.3.8}$$

so that, by selecting $i=(\sigma_0,\ldots,\sigma_{N-1})$, $x_i = p_1\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$, $y_i = p_2\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$, then (3.3.8) and (3.3.2) imply the first of the (3.3.5).

The second of (3.3.5) is based on proposition (3.2.1): if $s(\underline{p}_1)=s_1 \le s(\underline{p}_2)=s_2$ and if $\mathcal{C}^1_{1,\varepsilon}(N)$ and $\mathcal{C}^2_{1,\varepsilon}(N)$ denote the sets of specifications of large probability with respect to $m_{\underline{p}_1}$ and $m_{\underline{p}_2}$ we see that

e3.3.9
$$\begin{aligned}
|\,\mathcal{C}^1_{1,\varepsilon}(N)\cup\mathcal{C}^2_{1,\varepsilon}(N)\,| &\le e^{N(s_1+\varepsilon)}+e^{N(s_2+\varepsilon)} \le\\
&\le e^{N(s_2+\varepsilon)}(1+e^{-N(s_2-s_1)}),
\end{aligned}\tag{3.3.9}$$

that shows, since

e3.3.10
$$\sum_{\sigma_0\ldots\sigma_{N-1}\notin\mathcal{C}^1_{1,\varepsilon}(N)\cup\mathcal{C}^2_{1,\varepsilon}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le 2\varepsilon, \tag{3.3.10}$$

that $s = s(\underline{p}) \le s_2$.

But one cannot have $s < s_2$: if indeed $\overline{\mathcal{C}}_\varepsilon(N)$ was a set of large probability ($> 1-\varepsilon$) and $|\,\overline{\mathcal{C}}_\varepsilon(N)\,| < e^{N(s+\varepsilon)}$, for all $N \ge N_\varepsilon$, we should have

e3.3.11
$$\begin{aligned}&\sum_{\sigma_0\ldots\sigma_{N-1}\notin\overline{\mathcal{C}}_\varepsilon(N)} p_2\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} =\\ &= (1-a)^{-1}(1-a)\sum_{\sigma_0\ldots\sigma_{N-1}\notin\overline{\mathcal{C}}_\varepsilon(N)} p_2\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le\\ &\le (1-a)^{-1}\sum_{\sigma_0\ldots\sigma_{N-1}\notin\overline{\mathcal{C}}_\varepsilon(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le \varepsilon(1-a)^{-1},\end{aligned} \tag{3.3.11}$$

but this would contradict the fact that s_2 is the entropy of $\underline{p}_2$. ∎

Another simple but important property of $s(\underline{p})$ is the following one.

P3.3.2 **(3.3.2) Proposition:** (Average entropy as an infimum, semicontinuity) *Let $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ be an invariant distribution; one has*

e3.3.12
$$\begin{aligned}&\widetilde{s}(\underline{p}) \le \log(1+n), \qquad\qquad (3.3.12)\\ &\widetilde{s}(\underline{p}) = \inf_{N=2^k} -N^{-1}\sum_{\sigma_0\ldots\sigma_{N-1}} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix},\end{aligned}$$

where the infimum is taken over the integers k, by setting $N = 2^k$. Hence the average entropy is upper-semicontinuous, i.e. if $\underline{p}_n$ is a sequence of distributions which converges to a limit $\underline{p}_\infty$ in the topology of $M(\{0,\ldots,n\}^{\mathbb{Z}})$, i.e. in the sense that $p_n\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \xrightarrow[n\to\infty]{} p_\infty\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$ for all $N, \sigma_0, \sigma_1, \ldots, \sigma_N$, then

e3.3.13
$$\widetilde{s}(\underline{p}_\infty) \ge \limsup_{n\to\infty} \widetilde{s}(\underline{p}) \tag{3.3.13}$$

and $s(\mu)$ has a maximum on any compact set in $M(\{0,\ldots,n\}^{\mathbb{Z}})$.

Proof: The function $(x_1,\ldots,x_p) \to -\sum_i x_i \log x_i$, $0 \le x_1,\ldots,x_p \le 1$, $\sum_i x_i = 1$, has its maximum in $x_i = 1/p$ where its value is $\log p$. The sum (3.3.2) has precisely this form: this shows the first relation of (3.3.12).

To show the second relation in (3.3.12) let I and J be two sets of labels and let $(p_i)_{i\in I}$, $(p'_j)_{j\in J}$, $(p_{ij})_{ij\in I\times J}$ be three families of not negative numbers
N3.3.2 such that [2]

e3.3.14
$$\begin{aligned}&\sum_i p_i = \sum_j p'_j = \sum_{ij} p_{ij} = 1,\\ &p_i = \sum_j p_{ij}, \qquad p'_j = \sum_i p_{ij}.\end{aligned} \tag{3.3.14}$$

[2] We think here to $i = (\sigma_0\ldots\sigma_{N-1})$, $j = (\sigma'_0\ldots\sigma'_{N-1})$, $p_i = p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$, $p'_j = p\begin{pmatrix}0\ldots M-1\\ \sigma'_0\ldots\sigma'_{M-1}\end{pmatrix}$, $p_{ij} = p\begin{pmatrix}0\ldots N-1 & N\ldots N+M-1\\ \sigma_0\ldots\sigma_{N-1} & \sigma'_0\ldots\sigma'_{M-1}\end{pmatrix}$.

Then

e3.3.15
$$\begin{aligned}\sum_{ij} -p_{ij}\log p_{ij} &\le \sum_i -p_i \log p_i + \sum_j -p'_j \log p'_j,\\ \sum_{ij} -p_{ij}\log p_{ij} &\ge \sum_i -p_i\log p_i,\end{aligned}\tag{3.3.15}$$

because

e3.3.16
$$\begin{aligned}\sum_{ij} - p_{ij}\log p_{ij} &= \sum_i p_i \sum_j -(p_{ij}/p_i)\log p_{ij} =\\ &= \sum_i p_i \sum_j -(p_{ij}/p_i)(\log p_{ij}/p_i + \log p_i) =\\ &= \sum_i p_i \sum_j -(p_{ij}/p_i)\log(p_{ij}/p_i) + \sum_i -p_i\log p_i =\\ &= \sum_j \Big\{ \sum_i p_i(-p_{ij}/p_i)\log p_{ij}/p_i)\Big\} + \sum_i -p_i \log p_i\end{aligned}\tag{3.3.16}$$

and the term in curly bracket is ≥ 0 because $(p_{ij}/p_i) \le 1$; moreover by the convexity of $\xi \to -\log\xi$, it is bounded above by

e3.3.17
$$-\sum_j \big(\sum_i p_i(p_{ij}/p_i)\big)\log\big(\sum_i p_i(p_{ij}/p_i)\big) = -\sum_j p'_j \log p'_j. \tag{3.3.17}$$

Choosing p_i, p_{ij} as in footnote 2, and setting

e3.3.18
$$H_N(\underline{p}) = \sum_{\sigma_0\ldots\sigma_{N-1}} -p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}, \tag{3.3.18}$$

we see that the first of the (3.3.15) means

e3.3.19
$$H_{N+M}(\underline{p}) \le H_N(\underline{p}) + H_M(\underline{p}), \tag{3.3.19}$$

which implies that $N^{-1}H_N(\underline{p})$ is monotonic non-increasing on the sequence $N = 2^k$, $k = 0, 1, \ldots$ and, hence, the second of (3.3.12) follows. ■

The second of the (3.3.15) will be useful in the following and sometimes we shall refer directly to it without formulating it as a separate proposition.

The following definition is remarkable and natural.

D3.3.2 **(3.3.2) Definition:** (Average entropy of a dynamical system)
Let (Ω, S, μ) be an invertible dynamical system and let $\mathcal{P} = \{P_0, \ldots, P_n\}$ be a partition of Ω into μ–measurable sets. We define the average entropy of S with respect to $\mathcal{P}$ and μ *the following quantity:*

e3.3.20
$$\widetilde{s}(\mathcal{P}, S, \mu) = \widetilde{s}(\underline{p}_\mu), \tag{3.3.20}$$

where $\underline{p}_\mu \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ *is defined by*

e3.3.21
$$p_\mu \begin{pmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} = \mu\big(\bigcap_{k=0}^{N-1} S^{-k}P_{\sigma_k}\big). \tag{3.3.21}$$

We define the average entropy of S with respect to μ *the quantity*

e3.3.22
$$\widetilde{s}(S,\mu) = \sup_{\mathcal{P}} \widetilde{s}(\mathcal{P},S,\mu), \tag{3.3.22}$$

where the supremum is considered over all the finite partitions of Ω *into* μ*–measurable sets.*

Remarks: (1) Noting that $p_\mu \begin{pmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} = p_\mu \begin{pmatrix} -N+1\ldots 0 \\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$ we deduce from (3.3.2) that

e3.3.23
$$\widetilde{s}(\mathcal{P},S,\mu) = \widetilde{s}(\mathcal{P},S^{-1},\mu). \tag{3.3.23}$$

(2) Furthermore it is clear that $\widetilde{s}(S,\mu)=\widetilde{s}(S',\mu')$ if (Ω,S,μ) is isomorphic mod 0 to (Ω',S',μ'): *the average entropy of a metric dynamical system is an invariant under isomorphisms mod* 0. For this reason the average entropy is also called the *Kolmogorov–Sinai invariant.*

Problems for §3.3

Q3.3.1 **[3.3.1]:** Making use of proposition (3.3.2) and of problem [3.2.8] show that, given $\varepsilon>0$, one can find a partition $\mathcal{P}_\varepsilon$ of $\{0,1\}^{\mathbb{Z}}$ that has average entropy $<\varepsilon$ with respect to the action of the translation on the Bernoulli measure $B(1/2,1/2)$. (*Hint:* Construct a partition that interpolates between the partition $\mathcal{P}_0=\{\emptyset,\{0,1\}^{\mathbb{Z}}\}$ and $\mathcal{P}_1=\{C^0_0,C^0_1\}$ and estimate $\widetilde{s}(\mathcal{P}_\varepsilon,S,\mu)$ by means of (3.3.12) with $N=1$).

Q3.3.2 **[3.3.2]:** Consider the Bernoulli scheme $B(1/2,1/2)$ and compute the entropy of the partition $\{C^{01}_{00},C^{01}_{01},C^{01}_{10},C^{01}_{11}\}$.

Q3.3.3 **[3.3.3]:** By using the results of problem [2.4.6] show that if $\Delta=\{\omega\,|\,\omega\in\mathcal{M}_e(\{0,\ldots,n\}^{\mathbb{Z}},\tau),\, s(\omega)\in[\alpha,\beta]\}$, and if π is a Borel measure on $\mathcal{M}_e$ such that $\pi(\Delta)=1$, then the measure $m=\int_\Delta \omega\pi(d\omega)$ has entropy $\widetilde{s}(m)\in[\alpha,\beta]$. The same happens if $[\alpha,\beta]$ is replaced, in the definition of Δ, by $[\alpha,\beta),(\alpha,\beta],(\alpha,\beta)$. (*Hint:* Note that $\varepsilon(\Delta)$ is a Borel set and make use of problem [2.4.6].)

Q3.3.4 **[3.3.4]:** (*Affinity of average entropy for finite mixtures*)
By using problem [3.3.3] show that the average entropy s is affine with respect to the ergodic decompositions of measures $\underline{p}\in M(\{0,\ldots,n\}^{\mathbb{Z}})$, *i.e.* show that the affinity in proposition (3.3.1), for finite mixtures, implies via problem [3.3.3] and the Shannon–McMillan theorem, that if $\underline{p}\in M(\{0,\ldots,n\}^{\mathbb{Z}})$ and π is its ergodic decomposition, then

$$\widetilde{s}(\underline{p}) = \int_{M_e} \pi(d\omega)\,\widetilde{s}(\omega).$$

Q3.3.5 **[3.3.5]:** (*Affinity of the average entropy for arbitrary mixtures*)
If (Ω,S) is an invertible topological dynamical system with Ω metric and compact and if $\mu\in\mathcal{M}(\Omega,S)$ and π_μ is the ergodic decomposition on $\mathcal{M}_e(\Omega,S)$ of μ then

$$\widetilde{s}(S,\mu) = \int_{\mathcal{M}_e(\Omega,S)} \pi_\mu(d\omega)\,\widetilde{s}(\omega),$$

i.e. the average entropy is *affine* also for ergodic decompositions which are not finite.

Q3.3.6 **[3.3.6]:** Let ρ be a probability distribution on $\{0,\ldots n\}^N$. Construct the measure μ_0 on $\{0,\ldots,n\}^{\mathbb{Z}}$ by assigning independent probabilities q to the blocks of variables $(\sigma_i)_{i\in[kN,(k+1)N)}$, $k\in\mathbb{Z}$. Show that μ_0 is invariant with respect to the action of translations which are multiples of the N steps translation τ^N. Show that, for $N>1$,

$$\mu(E)=N^{-1}\sum_{j=0}^{N-1}\mu_0(\tau^j E)\qquad\text{for all } E\in\mathcal{B}(\{0,\ldots,n\}^{\mathbb{Z}})$$

defines a τ–ergodic and τ^N–mixing measure which is not τ–mixing.

Q3.3.7 **[3.3.7]:** (*Average entropy of periodic distributions*)
Compute the average entropy of the measure μ defined in problem [3.3.6] regarded as an invariant measure for the dynamical system $(\{0,\ldots,n\}^{\mathbb{Z}},\tau^N)$.

Q3.3.8 **[3.3.8]:** (*Approximability of measures by ergodic measures*)
If m is a shift invariant distribution on $\{0,\ldots,n\}^{\mathbb{Z}}$ and if $\varepsilon>0$ and M are given, there exists a probability distribution μ ergodic on $\{0,\ldots,n\}^{\mathbb{Z}}$ and such that

$$\sum_{\sigma_1\ldots\sigma_M}|\,m(C^{1\ldots M}_{\sigma_1\ldots\sigma_M})-\mu(C^{1\ldots M}_{\sigma_1\ldots\sigma_M})\,|<\varepsilon$$

(*Hint:* Consider $N\gg M$ and the distribution q on the sequences $\{0,\ldots,n\}^N$ defined by: $q(\sigma_1,\sigma_2,\ldots,\sigma_N)=m(C^{1,\ldots,N}_{\sigma_1,\ldots,\sigma_N})$; construct μ, starting from q, m as in problem [3.3.6]).

Q3.3.9 **[3.3.9]:** (*Approximability in distribution and entropy*)
Show that, as consequence of Shannon-McMillan theorem, every ergodic $m\in M_e(\{0,\ldots,n\}^{\mathbb{Z}})$ can be approximated in distribution and entropy by a measure μ which is ergodic and "of finite type", *i.e.* built as in problem [3.3.6]. By approximation in distribution and entropy one means that given $\varepsilon>0$ and $M>0$ there exists a μ of finite type for which

$$\sum_{\sigma_1\ldots\sigma_M}|\,m(C^{1\ldots M}_{\sigma_1\ldots\sigma_M})-\mu(C^{1\ldots M}_{\sigma_1\ldots\sigma_M})\,|<\varepsilon,\qquad|\,s(\mu)-s(m)\,|<\varepsilon.$$

(*Hint:* Make use of problems [3.3.6], [3.3.7] and [3.3.8], and of (i) in proposition (3.2.1)).

Bibliographical note to §3.3

The properties discussed in this section are well known, see for instance p. 178 in [Ru69]. There are various other properties of entropy and mainly its extensions to "non-commutative cases" that are less simple and at times quite deep; see [We79], for a review.

§3.4 Further properties of the average entropy. Generator theorem

In this section we mainly discuss definition (3.3.2) and certain simplifications in the evaluation of the extremum in (3.3.22).

D3.4.1 **(3.4.1) Definition:** (Generating partition)
Let $\mathcal{B}$ *be a* σ*–algebra in a space* Ω *and* $\mathcal{P} = (\{P_0, \ldots, P_n\}$, $\mathcal{Q} = \{Q_0, \ldots, Q_m\}$ *be two* $\mathcal{B}$*-measurable partitions of* Ω*. Define the partition* $\mathcal{P} \vee \mathcal{Q}$, generated by $\mathcal{P}$ and $\mathcal{Q}$*, to be the partition whose atoms are*

e3.4.1
$$R_{\sigma\sigma'} = P_\sigma \cap Q_{\sigma'} \quad \sigma \in \{0, \ldots, n\},\ \sigma' \in \{0, \ldots, m\}. \tag{3.4.1}$$

If μ *is a probability measure on* $\mathcal{B}$ *define*

e3.4.2
$$H(\mathcal{P}, \mu) = \sum_{\sigma=0}^{n} -\mu(P_\sigma) \log \mu(P_\sigma). \tag{3.4.2}$$

If (Ω, S, μ) *is an invertible metric system with* μ *defined on* $\mathcal{B}$ *we say that* $\mathcal{P}$ *is* μ-S–generating *if the smallest* σ*–algebra that contains the sets of the partitions* $S^k\mathcal{P}$*, for all* $k \in \mathbb{Z}$*, coincides* μ*-mod* 0 *with* $\mathcal{B}$*. When confusion does not arise we shall simply say that* $\mathcal{P}$ *is* S*–generating.*

Remarks: (1) By using the definition given in (3.3.2), the quantity $\widetilde{s}(\mathcal{P}, S, \mu)$ can also be rewritten as (cf. also (3.3.18))

e3.4.3
$$\widetilde{s}(\mathcal{P}, S, \mu) = \lim_{N\to\infty} N^{-1} H(\mathcal{P} \vee S^{-1}\mathcal{P} \vee \ldots \vee S^{-(N-1)}\mathcal{P}, \mu). \tag{3.4.3}$$

(2) The identity (3.4.3) implies that for all h, k integer (positive, zero or negative), with $h \le k$, one has

e3.4.4
$$\widetilde{s}(S^h\mathcal{P} \vee \ldots \vee S^k\mathcal{P}, S, \mu) = \widetilde{s}(\mathcal{P}, S, \mu). \tag{3.4.4}$$

(3) From general measure theory it follows that if μ is isomorphic mod 0 to the Lebesgue measure on $[0, 1]$ then a necessary and sufficient condition in order that $\mathcal{P}$ be S–generating is that $\mathcal{P}$ is S*–separating mod* 0, *i.e.* that there exists $N \in \mathcal{B}$, $\mu(N) = 0$, such that if $x, y \notin N$ and the $(\mathcal{P}, S)$–histories of x and y coincide then x and y coincide too.

Hence in dynamical systems (Ω, S, μ) in which (Ω, S) is a topological dynamical system and μ a Borel measure every S–separating partition is generating.[1] N3.4.1

(4) From general measure theory it follows that in order that $\mathcal{P}$ be S–generating it must happen that, given $\varepsilon > 0$ and $E \in \mathcal{B}$, there exists N_ε such that the partition $\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$ is "fine enough" so that it is possible, by taking suitable unions of its atoms, to construct a set E_ε whose symmetric difference from E, $(E\Delta E_\varepsilon = (E\backslash E_\varepsilon) \cup (E_\varepsilon \backslash E)$, is small

e3.4.5
$$\mu(E\Delta E_\varepsilon) < \varepsilon. \tag{3.4.5}$$

(5) From remark (4) it follows that if $\mathcal{P}$ is S–generating and if $\mathcal{Q} = \{Q_0, \ldots, Q_m\}$ is an arbitrary μ–measurable partition, given $\varepsilon > 0$ there

[1] Every Borel measure on a complete and separable metric space is isomorphic mod 0 to the sum of the Lebesgue measure on an interval and a denumerable sum of Dirac measures, cf. problem [3.2.9] and [Pa67].

exists N_ε such that by suitably collecting the atoms of $\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$ into $(m+1)$ groups one can form a partition $\mathcal{Q}^\varepsilon = \{Q_0^\varepsilon, \ldots, Q_m^\varepsilon\}$ such that

e3.4.6
$$d(\mathcal{Q}, \mathcal{Q}^\varepsilon) = \sum_{i=0}^{n} \mu(Q_i \Delta Q_i^\varepsilon) < \varepsilon, \tag{3.4.6}$$

where, given in general two partitions $\mathcal{Q} = \{Q_0, \ldots, Q_m\}$ and $\mathcal{Q}' = \{Q_0', \ldots, Q_m'\}$ with an equal number of atoms, we define

e3.4.7
$$d(\mathcal{Q}, \mathcal{Q}') = \sum_{j=0}^{n} \mu(Q_i \Delta Q_i'). \tag{3.4.7}$$

The interest of the above observations and their relevance for the problem of the actual computation of the extreme in (3.3.22) lies in the corollary to the following proposition.

P3.4.1 **(3.4.1) Proposition:** (Continuity of the average entropy of a partition) *Let (Ω, S, μ) be an invertible metric dynamical system.*
(i) If $\mathcal{P} = \{P_0, \ldots, P_n\}$, $\mathcal{Q} = \{Q_0, \ldots, Q_m\}$ are two μ–measurable partitions of Ω, then

e3.4.8
$$\widetilde{s}(\mathcal{P}, S, \mu) \le \widetilde{s}(\mathcal{P} \vee \mathcal{Q}, S, \mu) \le \widetilde{s}(\mathcal{P}, S, \mu) + \widetilde{s}(\mathcal{Q}, S, \mu). \tag{3.4.8}$$

(ii) If $\mathcal{P} = \{P_0, \ldots, P_n\}$ and $\mathcal{Q} = \{Q_0, \ldots, Q_n\}$ are two μ–measurable partitions and if $d(\mathcal{P}, \mathcal{Q}) = \sum_{i=0}^{n}(P_i \Delta Q_i) = \varepsilon < (n+1)/(n+2)$, one has

e3.4.9
$$|\widetilde{s}(\mathcal{P}, S, \mu) - \widetilde{s}(\mathcal{Q}, S, \mu)| \le \varepsilon \log(n+1) - \varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon). \tag{3.4.9}$$

Remark: The statement (ii), *Sinai's theorem*, gives continuity in $\mathcal{P}$ at fixed number, n, of atoms.

Proof: Let $i = (\sigma_0 \ldots \sigma_{N-1})$, $j = (\lambda_0 \ldots \lambda_{N-1})$, $\sigma_k \in \{0, \ldots, n\}$, $\lambda_k \in \{0, \ldots, m\}$ and

e3.4.10
$$\begin{aligned} &p_i = \mu\Big(\bigcap_{k=0}^{N-1} S^{-k} P_{\sigma_k}\Big), \qquad p_{ij} = \mu\Big(\bigcap_{k=0}^{N-1} S^{-k}(P_{\sigma_k} \cap Q_{\lambda_k})\Big),\\ &p_j' = \mu\Big(\bigcap_{k=0}^{N-1} S^{-k} Q_{\lambda_k}\Big); \end{aligned} \tag{3.4.10}$$

one has $\sum_i p_i = \sum_{ij} p_{ij} = 1$, $\sum_j p_{ij} = p_i$, $\sum_i p_{ij} = p_j'$. Then (3.4.8) is derived from the relation between the approximants $\sum_i(-p_i \log p_i)$, $\sum_{ij}(-p_{ij}\log p_{ij})$, $\sum_i(-p_i \log p_i + \sum_j -p_j' \log p_j')$ obtained in (3.3.15) from (3.3.14).

To show (ii) let $\mathcal{R} = \{R_0, \ldots, R_n, R_{n+1}\}$ be the partition of Ω into $n+2$ sets defined by

e3.4.11
$$\begin{aligned} R_0 &= P_0 \Delta Q_0, \\ R_i &= (P_i \Delta Q_i) / \bigcup_{j=0}^{i-1} (P_j \Delta Q_j), \qquad i = 1, \ldots, n, \\ R_{n+1} &= \Omega / \bigcup_{i=0}^{n} R_i. \end{aligned} \tag{3.4.11}$$

We have $X \stackrel{def}{=} \mu(R_{n+1}) \geq 1 - \varepsilon$ and $\sum_{i=0}^{n} \mu(R_i) = 1 - X \leq \varepsilon$. Furthermore (3.4.8) implies

e3.4.12
$$\widetilde{s}(\mathcal{P} \vee \mathcal{R}, S, \mu) \leq \widetilde{s}(\mathcal{P}, S, \mu) + \widetilde{s}(\mathcal{R}, S, \mu), \tag{3.4.12}$$

and the similar relation with $\mathcal{Q}$ instead of $\mathcal{P}$. The relation $(\mathcal{P} \vee \mathcal{R}) = (\mathcal{Q} \vee \mathcal{R})$ (that follows immediately from the definitions) implies together with (3.4.12)

e3.4.13
$$|\widetilde{s}(\mathcal{P}, S, \mu) - \widetilde{s}(\mathcal{Q}, S, \mu)| \leq \widetilde{s}(\mathcal{R}, S, \mu). \tag{3.4.13}$$

But from the second of (3.3.12) with $k = 0$ it follows

e3.4.14
$$\widetilde{s}(\mathcal{R}, S, \mu) \leq \sum_{\sigma=0}^{n+1} -\mu(R_\sigma) \log \mu(R_\sigma), \tag{3.4.14}$$

which implies

e3.4.15
$$\begin{aligned} \widetilde{s}(\mathcal{R}, S, \mu) &\leq -X \log X + \sum_{\sigma=0}^{n} -\mu(R_\sigma) \log \mu(R_\sigma) \equiv \\ &\equiv -X \log X - (1-X) \log(1-X) + (1-X) \sum_{\sigma=0}^{n} -\frac{\mu(R_\sigma)}{(1-X)} \log \frac{\mu(R_\sigma)}{(1-X)}; \end{aligned} \tag{3.4.15}$$

therefore the sum can be bounded above by $\log(n+1)$ because of the identity $\sum_{\sigma=0}^{n} \frac{\mu(R_\sigma)}{1-X} = 1$. Hence

e3.4.16
$$\widetilde{s}(\mathcal{R}, S, \mu) \leq -X \log X - (1-X) \log(1-X) + (1-X) \log(n+1), \tag{3.4.16}$$

from which (3.4.9) follows because the function in the r.h.s. of (3.4.16) is monotonic decreasing between $(1 - (n+1)/(n+2))$ and 1, and one has $X \geq 1 - \varepsilon$. ■

C3.4.1 **(3.4.1) Corollary:** (Generator theorem)
Let (Ω, S, μ) be an invertible metric dynamical system. If $\mathcal{P} = \{P_0, \ldots, P_n\}$, $\mathcal{Q} = \{Q_0, \ldots, Q_m\}$ are two μ–measurable partitions and $\mathcal{P}$ is S–generating, then

e3.4.17
$$\widetilde{s}(\mathcal{Q}, S, \mu) \leq \widetilde{s}(\mathcal{P}, S, \mu) = \widetilde{s}(S, \mu) \tag{3.4.17}$$

Remark: The above theorem is due to Sinai.

Proof: Let $\varepsilon > 0$ and let $\mathcal{Q}^\varepsilon$ be a partition obtained by forming unions of atoms of $\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$ and such that $d(\mathcal{Q},\mathcal{Q}^\varepsilon) < \varepsilon$, cf. remark (5) to definition (3.4.1). Then, by proposition (3.4.1)

e3.4.18
$$|\widetilde{s}(\mathcal{Q},S,\mu) - \widetilde{s}(\mathcal{Q}^\varepsilon,S,\mu)| \xrightarrow[\varepsilon\to 0]{} 0. \tag{3.4.18}$$

But it is clear that $\mathcal{Q}^\varepsilon$ is less fine than $\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$ (and therefore there is a partition $\mathcal{R}$ of Ω such that $\mathcal{Q}^\varepsilon \vee \mathcal{R} = \vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$); hence by (3.4.8) and (3.4.4)

e3.4.19
$$\widetilde{s}(\mathcal{Q}^\varepsilon,S,\mu) \le \widetilde{s}\Big(\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P},S,\mu\Big) = \widetilde{s}(\mathcal{P},S,\mu), \tag{3.4.19}$$

so that the first of (3.4.17) is proved. The second follows by the arbitrariness of $\mathcal{Q}$ and the definition in (3.3.22). ■

C3.4.2 **(3.4.2) Corollary:** *Let (Ω,S,μ) be an invertible metric system mod 0 with μ defined on a σ–algebra $\mathcal{B}$. Let $\mathcal{P}_1$, $\mathcal{P}_2,\ldots$ be a sequence of μ–measurable partitions such that $\mathcal{P}_{n+1}$ refines $\mathcal{P}_n$ (i.e. such that the atoms of $\mathcal{P}_n$ are unions of atoms of $\mathcal{P}_{n+1}$) and such that the smallest σ–algebra that contains $\mathcal{P}_n$ and all its images under S, for all $n = 1,2\ldots$, is $\mathcal{B}$ (*a generating sequence of partitions*). One has*

e3.4.20
$$\widetilde{s}(S,\mu) = \lim_{n\to\infty} \widetilde{s}(\mathcal{P}_n,S,\mu). \tag{3.4.20}$$

The proof will be left to the reader. Finally an application.

P3.4.2 **(3.4.2) Proposition:** (Entropy bound for smooth measures)
Let Ω be a compact Riemannian manifold of class C^∞ and dimension r, S be a C^∞ diffeomorphism of Ω, and μ be an S–invariant Borel measure equivalent to the volume measure μ_0 on Ω (i.e. let $\mu = \rho\mu_0$ with $\rho^{\pm1} \in L_1(\mu_0)$). The average entropy of S with respcet to μ can be bounded as

e3.4.21
$$\widetilde{s}(S,\mu) < r\,\log\lambda, \tag{3.4.21}$$

in terms of the largest expansion coefficient λ of the line elements of Ω under the action of $S^{\pm1}$.

Remark: This is, essentially, again Kouchnirenko's theorem, cf. proposition (3.1.2): this time it is formulated on the average entropy and without the hypothesis of ergodicity of μ (so that the average entropy cannot be identified with the entropy).

Proof: Since Ω is locally diffeomorphic to $\mathbb{R}^r$ it is clear that there exists a sequence of partitions with sets with a piecewise C^∞ boundary and which verify the hypothesis of corollary (3.4.2). We can in fact suppose that the atoms of such partitions have always diameter less than a prefixed $\delta > 0$. We shall fix δ so that the isoperimetric inequality holds for every C^∞–regular set P with diameter diam $(P) < \delta$

e3.4.22
$$|\mu(P)| \le \Gamma |\partial P|^{r/(r-1)}, \tag{3.4.22}$$

where $|\partial P| =$ (area of the surface ∂P), and Γ is a suitable P–independent constant.

By corollary (3.4.2) it will suffice to show that for an arbitrary partition $\mathcal{P}$, C^∞–regular, with atoms of diameter $\le \delta$ one has $s(\mathcal{P}, S, \mu) \le r \log \lambda$.

Fixed $\mathcal{P}$ and $\eta > 0$ and proceeding as in proposition (3.1.2) we split the specifications $\sigma_0, \ldots, \sigma_{N-1}$ of length N, $\sigma_i \in \{0, \ldots, n\}$ into two classes:

e3.4.23
$$\begin{aligned} &\mathcal{C}_1(N) = \big\{\sigma_0, \ldots, \sigma_{N-1} \big| \mu\big(\bigcap_{k=0}^{N-1} S^{-k} P_{\sigma_k}\big) > e^{-N\eta} \lambda^{-rN}\big\}, \\ &\mathcal{C}_2(N) = \big\{0, \ldots, n\big\}^N \backslash \mathcal{C}_1(N), \end{aligned} \tag{3.4.23}$$

and, as in the case of the mentioned proposition, we deduce that if

e3.4.24
$$p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} = \mu\big(\bigcap_{k=0}^{N-1} S^{-k} P_{\sigma_k}\big) \tag{3.4.24}$$

one has, for a suitable $C > 0$,

e3.4.25
$$X \stackrel{def}{=} \sum_{\sigma_0 \ldots \sigma_{N-1} \in \mathcal{C}_2(N)} p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \le C e^{-N\eta}. \tag{3.4.25}$$

Hence. having set $j = (\sigma_0, \ldots, \sigma_{N-1})$, $p_j = p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix}$,

e3.4.26
$$\begin{aligned} &\sum_j -p_j \log p_j = \sum_{j \in \mathcal{C}_1(N)} -p_j \log p_j + \sum_{j \in \mathcal{C}_2(N)} -p_j \log p_j \le \\ &\le (r \log \lambda + \eta) N + X \sum_{j \in \mathcal{C}_2(N)} -(p_j/X) \log p_j = \\ &= (r \log \lambda + \eta) N - X \log X + X \sum_{j \in \mathcal{C}_2(N)} (-p_j/X) \log(p_j/X) = \\ &= (r \log \lambda + \eta) N - X \log X + X N \log(n+1), \end{aligned} \tag{3.4.26}$$

because {number of elements in $\mathcal{C}_2(N)$} $\le (n+1)^N$. Dividing the (3.4.26) by N and passing to the limit as $N \to \infty$ the terms containing X tend to zero and one finds $s(\mathcal{P}, S, \mu) \le \eta + r \log \lambda$, for every $\eta > 0$. ■

Problems for §3.4 (*Complements to Shannon–McMillan's theorem*)

Q3.4.1 **[3.4.1]:** (*Average entropy of a Bernoulli scheme*)
Consider the Bernoulli scheme on $\Omega = \{0, \ldots, n\}^{\mathbb{Z}}$ that associates with the symbols

the probabilities $\pi_0, \ldots, \pi_n$, $\sum_i \pi_i = 1$. Denoting by m the corresponding probability measure on Ω consider the system (Ω, τ, m) and show that its entropy $s = s(\tau, m)$ and its average entropy $\widetilde{s} = \widetilde{s}(\tau, m)$ satisfy $s = \widetilde{s} = -\sum_i \pi_i \log \pi_i$. (*Hint:* Recall the definition in (3.3.22), and use problem [3.1.8] and the Shannon–McMillan theorem.)

Q3.4.2 **[3.4.2]:** Show the existence of Bernoulli schemes (with infinitely many different symbols) with infinite average entropy. (*Hint:* Choose a sequence $\{a_n\}_{n=0}^{\infty}$ such that $\sum_{n=1}^{\infty} a_n < \infty$ and $\sum_{n=1}^{\infty} a_n \log a_n < \infty$, and normalize to 1 the first series; take for instance $a_n = 1/n \log^2 n$ for $n \geq 2$.)

Q3.4.3 **[3.4.3]:** (*Average entropy of a Markov chain*)
Compute the average entropy of the dynamical system $(\{0,1\}^{\mathbb{Z}}, \tau, m)$ where m is the distribution built in [2.3.8]. Show that $\widetilde{s}(\tau, m) = -\sum_{\sigma,\sigma'} \pi^*_\sigma \pi_{\sigma'} \frac{T_{\sigma\sigma'}}{\lambda} \log \frac{T_{\sigma\sigma'}}{\lambda}$.

Q3.4.4 **[3.4.4]:** Estimate the average entropy of the system of the example (1.2.5).

Q3.4.5 **[3.4.5]:** Show that two Bernoulli schemes with different entropy cannot be isomorphic mod 0.

Q3.4.6 **[3.4.6]:** (*A non–generating partition with maximal entropy*)
Consider the Markov process (cf. problem [2.3.8]) with transition matrix $T_{\sigma\sigma'} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ and show that it admits a *non-generating* partition $\mathcal{Q}$ with largest entropy. (*Hint:* $Q_0 = \{\underline{\sigma} | \sigma_0 = \sigma_1\}$, $Q_1 = \{\underline{\sigma} | \sigma_0 \neq \sigma_1\}$, $s(\mathcal{Q}, \tau) = \log 2$.)

Q3.4.7 **[3.4.7]:** Let μ_n be a sequence of invariant distributions for a topological dynamical system (Ω, S) and suppose that there is a partition $\mathcal{P}$ which is S–generating for all (Ω, S, μ_n). Then if μ_n converges weakly to μ, *i.e.* $\mu_n(f) \xrightarrow[n\to\infty]{} \mu(f)$ for all continuous functions f, the average entropies of μ_n verify $\limsup_{n\to\infty} \widetilde{s}(S, \mu_n) \leq \widetilde{s}(S, \mu)$. (*Hint:* See (3.3.13).)

Q3.4.8 **[3.4.8]:** (*Factors of arbitrary entropy*)
Show that given a dynamical system (Ω, S, μ) with a given partition $\mathcal{P}$, such that the measure μ on Ω is isomorphic mod 0 to the Lebesgue measure on $[0,1]$, there exists a partition $\mathcal{P}'$ of Ω such that $s(\mathcal{P}', S) = a\, s(\mathcal{P}, S)$ with a arbitrarily fixed in $[0,1]$. The dynamical system (Ω, S, μ') where μ' is the restriction of μ to the σ–algebra generated by $\mathcal{P}'$ is called a *factor* of the original dynamical system. (*Hint:* Use Sinai's theorem and the existence of a "continuous" family interpolating between the trivial partition and $\mathcal{P}$; proceed as in problem [3.2.8]). To appreciate the generality of this property note its relation with problems [3.2.1], [3.2.11] and [3.2.12].

Q3.4.9 **[3.4.9]:** (*Stacks for mixing systems*)
Consider a system $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, \mu)$ with μ mixing and non-atomic (hence such that $\sup \mu(C^{0\ldots k}_{\sigma_0\ldots\sigma_k}) \xrightarrow[k\to\infty]{} 0$) and, for simplicity, assume $\mu(C^{0\ldots k}_{\sigma_0\ldots\sigma_k}) > 0$ for all $\sigma_0, \ldots, \sigma_k$. Show that, given $\varepsilon > 0$ and $N > 0$ integer, there exists a Borel set F such that the sets F, $\tau F, \ldots, \tau^{N-1} F$ are pairwise disjoint and $\mu\Big(\cup_{i=0}^{N-1} \tau^i F\Big) \geq 1 - \varepsilon$. (*Hint:* : Let $M \gg M' \gg N$ and choose M' such that sup $\mu(C^{0\ldots M'}_{\sigma_0\ldots\sigma_{M'}}) < \varepsilon/4N$. Consider the string with M' elements $00000\ldots001$ and call $\widetilde{F}$ the union of the cylinders with base $(kN, \ldots, kN + M')$ and specification $00000\ldots001$ for $k = 0, 1, 2, \ldots, [M/N]$ and $F = \widetilde{F}/\cup_{i=1}^{N-1} \tau^i \widetilde{F}$ (*i.e.* F is the set of sequences for $k = 0, 1, 2, \ldots, [M/N]$ (ie F is the set of sequences containing the string $00000\ldots001$ with the first zero in position 0 or N or $2N$, *etc.*, for the first time between 0 and $M-1$). It is then clear that $\tau^i F \cap \tau^j F = \emptyset$ $\forall 0 \leq i \neq j \leq N-1$, and $\{0, \ldots, n\}^{\mathbb{Z}}/\cup \tau^i F$ contain two types of points: those sequences $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ that never contain between 0 and $M-1$ the string $00000\ldots001$ and those that do contain it between 0 and $N-1$ The set of the points of the second type has measure lower than $(\varepsilon/4N)N = \varepsilon/4$ and the set of the first type has infinitesimal measure as $M \to \infty$.)

Q3.4.10 **[3.4.10]:** (*Rohlin's stack*)
Under the hypothesis of problem [3.4.9] let $\mathcal{Q} = (Q_0, \ldots, Q_k)$ be a Borel partition of

$\{0,\ldots,n\}^{\mathbb{Z}}$. It is possible, given N and $\varepsilon > 0$, to find F so that $\mu(Q_i \cap F)/\mu(F) = \mu(Q_p)$ 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 result in problem [3.4.9] and verifying 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

$$\lim_{t\to\infty} \sum_{i=1}^{k} \left| \frac{\mu(Q_i \cap F_t)}{\mu(F_t)} - \mu(Q_i) \right| = \lim_{t\to\infty} \eta_t = 0$$

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. problem [3.2.12]) and regarding in this way $Q_0,\ldots,Q_k, F_t$ 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 $\mu(Q_i \cap (F_t \setminus \Delta)) = \mu(Q_i)\mu(F_t \setminus \Delta)$ without deteriorating the bound on the measure of $\cup_{i=0}^{N-1} \tau^i (F_t \setminus \Delta))$, *i.e.* keeping it larger than $1 - \varepsilon$).

Remark: This statement (*Rohlin's stack theorem*) does not require the hypothesis of mixing: ergodicity of μ suffices (the same can be said also for the result of the preceding problem [3.4.9]); however the proof is, in the latter cases, somewhat more elaborate.

The following problems provide a guided proof of the statement that two mixing shifts of equal entropy contain copies of each other (Sinai).

Q3.4.11 **[3.4.11]:** 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).

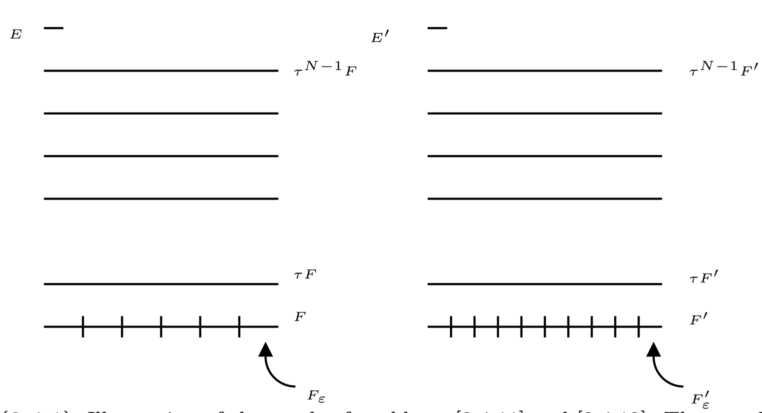
Given $\varepsilon > 0$ and u integer > 0 consider the partitions $\mathcal{P}_N = \vee_0^{N-1} \tau^i \mathcal{P}$ and $\mathcal{P}'_N = \vee_0^{N-1} \tau^i \mathcal{P}'$ and choose $N > N(u,\varepsilon)$, where $N(u,\varepsilon)$ is such that for $N > N(u,\varepsilon)$ the properties stated in problem [3.2.1] hold. Let $F \subset \{0,\ldots,n\}^{\mathbb{Z}}$ and $F' \subset \{0,\ldots,n'\}^{\mathbb{Z}}$ be two sets for which (cf. problem [3.4.10])

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

and simultaneously $\tau^i F \cap \tau^j F = \emptyset = \tau^i F' \cap \tau^j F'$, for all $i \ne j$, $i,j = 0,\ldots,N-1$. Represent F and F' as two intervals (see Fig. (3.4.1)), and represent as intervals also the sets

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

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

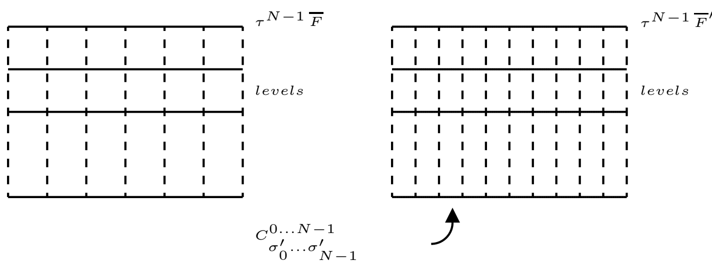


F3.4.1 **Fig.(3.4.1)** Illustration of the result of problems [3.4.11] and [3.4.12]. The sets F_ε and F'_ε are defined in problem [3.4.12]. The interval F, base of the stack, is divided into smaller intervals representing the sets $F \cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ with $\sigma_0 \ldots, \sigma_{N-1}$ chosen in the large frequency collection $\mathcal{C}_{1,\varepsilon,u}(N)$, except the rigthmost interval, denoted F_ε, which represents the intersection of F with $\cup C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ where the union is over the cylinders with $\sigma_0, \ldots, \sigma_{N-1}$ in the collection $\mathcal{C}_{2,\varepsilon,u}$ of rare strings.

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.

Q3.4.12 **[3.4.12]:** In the situation of problem [3.4.11] draw as segments the several elements of the partition induced by $\mathcal{Q} \equiv \mathcal{P}_N$ on F (cf. Fig. (3.4.1)): $(Q \cap F)_{Q\in\mathcal{Q}}$; and represent as an interval also the set $F_\varepsilon = \cup F \cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ where the union is over the choices of $(\sigma_0, \ldots, \sigma_{N-1})$ in the collection $\mathcal{C}_{2,\varepsilon,u}(N)$ introduced in [3.2.1]. Perform the same construction over the stack relative to F'. Check that $\mu(F_\varepsilon) \le \varepsilon\mu(F)$, $\mu'(F'_\varepsilon) \le \varepsilon\mu'(F')$.

Q3.4.13 **[3.4.13]:** In the situation of problems [3.4.11] and [3.4.12] set $\overline{F} = F/F_\varepsilon$ and $\overline{F}' = F'/F'_\varepsilon$: such sets are split into disjoint parts by the partitions $\overline{F} \cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ and $\overline{F}' \cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ 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. problems [3.4.12] and [3.2.1]). Fix also $N \gg u$. We shall call "level" of the stack every image $\tau^j(\overline{F} \cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}})$, 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. (3.4.2).



F3.4.2 **Fig.(3.4.2)** Illustration of the stack levels of the construction of problem [3.4.13].

Remark that by the Shannon–McMillan theorem, if $2\varepsilon < s(\mu') - s(\mu)$ (where ε is the same as in proposition (3.2.1)) and N is large enough, the "columns of levels" with base F' are much more numerous of those with base F (the number of columns is essentially given by (3.2.4)).

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 $\cup_{j=0}^{N-1} \tau^j\, \overline{F}$ 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 $\cup_{i=0}^{N-1} \tau^i\, \overline{F}$ to the partition into levels and, therefore, we reconstruct from it, by means of operations of unions of atoms, the partition $\mathcal{P}$ on $\cup_{i=0}^{N-1} \tau^i\, \overline{F}$, *etc.*)

Q3.4.14 **[3.4.14]:** Show that the partition $\widetilde{\mathcal{P}}$ built via the procedure illustrated in problems

[3.4.9] to [3.4.13], is such that

$$\begin{aligned}&\sum_{\sigma'_0\ldots\sigma'_{n-1}} |\mu'(C^{0\quad u-1}_{\sigma'_0\ldots\sigma'_{n-1}}) - \mu(P^{0\ldots u-1}_{\sigma'_0\ldots\sigma'_{n-1}})| \xrightarrow[N\to\infty,\varepsilon\to0]{} 0\\ &|s(\widetilde{\mathcal{P}}',\tau) - s(\mathcal{P},\tau)| \xrightarrow[\varepsilon\to0]{} 0\end{aligned}$$

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 strengthened form of the Shannon–McMillan theorem in problems [3.2.1], [3.2.3] 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 $\mu'(C^{0\ldots u-1}_{\sigma'_0\ldots\sigma'_{n-1}})$ in the sequences of $\mathcal{C}_{1,\varepsilon,u}(N)$.)

Q3.4.15 **[3.4.15]:** (*Copying a dynamical system into another*)
Deduce from the results of problem [3.4.14] that, under the same hypotheses of [3.4.11] 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 $\widetilde{\widetilde{\mathcal{P}}}'$ such that $s(\widetilde{\widetilde{\mathcal{P}}}',\tau) = s - 2\varepsilon$. Apply then again the construction of the [3.4.11], [3.4.12], [3.4.13] replacing $\mathcal{P}'$ with $\widetilde{\widetilde{\mathcal{P}}}'$ *etc*).

Bibliographical note §3.4

The generator theorem of Sinai is discussed here following Appendix 19, p. 163, of [AA69]. The problems on Rohlin's stack and its applications are drawn from Ornstein's theory of Bernoulli shifts isomorphisms, [Or74].