Journal of Statistical Physics, Vol. 90, Nos. 3/4, 1998

Relative Entropy and Identification of Gibbs Measures in Dynamical Systems

J.-R. Chazottes,[1] E. Floriani,[1] and R. Lima[1]

Received June 27, 1997; final October 31, 1997

In this work we explore the idea of using the relative entropy of ergodic measures for the identification of Gibbs measures in dynamical systems. The question we face is how to estimate the thermodynamic potential (together with a grammar) from a sample produced by the corresponding Gibbs state.

KEY WORDS: Relative entropy; thermodynamic formalism; Gibbs measures; grammars.

1. INTRODUCTION

There are many situations where the investigation of properties of nonlinear dynamics is simplified or even made possible by the use of symbolic dynamics. This is typically true in cases where a complete mathematical foundation of the method can be achieved by defining a coding map that "relates the dynamical system, topologically and measure theoretically, to shift spaces" (R. Bowen[(1)]). It is also true when the time series corresponding to some experimental measurement are encoded, because of some technical constraints (e.g., precision, frequency acquisition), as symbolic sequences which can be used, by means of some hypothesis, to explore the underlying dynamical system.[(2, 3)] This is mainly the situation we have in mind to motivate the present work.

In both cases we are faced to two complementary issues. The first one, which is of topological nature, is the possible existence of a grammar, a set of restrictive rules that allows only a subset of sequences to appear as orbits

[1] Centre de Physique Théorique C.N.R.S, Marseille, France, and Centro Internacional de Ciencias A.C., Campus Universitario UAEM-UNAM Chamilpa, Cuernavaca-Morelos, Mexico.

of the dynamical system after the encoding operation has been performed. The second issue is of measure theoretical nature. Since the type of allowed blocks of symbols in a sequence is fixed by the grammar rules, what remains to know is the mean return time of each of these blocks inside the sequence, or equivalently, if we suppose ergodicity, their asymptotic distribution. In other words we need to supply our symbolic dynamical system with an ergodic measure.

For a large class of systems, ergodic measures can be constructed by a powerful tool, the so called Gibbs ansatz.[1,4] These measures are interesting in several aspects. First, they arise as solutions of a variational principle which does not make direct reference to the dynamics. Moreover, they can be built through a spectral construction that shows how the dynamics weights the phase space in a cascade tree, on smaller and smaller scales. Finally, and it is the most important for us, the asymptotic frequencies of the blocks of symbols are approximately, in a sense to become clear below, given by a simple function of the block, the energy, which in turn can be computed using a potential. Since the last is a function defined on the configuration space, we can hope to have in this way a concise and tractable description of the recurrence times and the asymptotic distribution of the blocks associated with the different parts of phase space.

In ref. 5, P. Collet, A. Galves and A. Lopes addressed the question of how to identify the grammar matrix, being given a sample sequence produced by a Gibbs measure whose potential is known. In that article, the authors defined two selection procedures, the first based on a maximum likelihood and the second on a minimum entropy principle, both leading to a selection of the original grammar in the limit of diverging observation time.

In the present work we first address, in a sense, the complementary issue, that is to say, how to identify a potential starting from a sample sequence generated by an unknown Gibbs measure, when the grammar is known. Notice that our problem is related to the more general context of statistical inference methods.[6–8] See also ref. 9 for a recent paper with applications and bibliography. In our case, as we will discuss in the next section, what can be identified is the equivalence class of the original potential. Our approach is similar to that of ref. 5, as we take advantage in a fundamental way of the thermodynamic formalism for dynamical systems. Of course, in order to do so, we need to restrict ourselves to the class of ergodic measures represented by Gibbs measures. A different point of view is adopted in ref. 10, where the authors propose a universal scheme which applies to a class of stochastic processes, and converges in a precise metric to the sampled process when the observation time diverges, without referring to a particular class of ergodic measures.

We shall see that the two strategies we propose to select the original potential will exclude all the "wrong" potentials. However, the observation time that is necessary to exclude each one of these potentials depends on it, and therefore, in general, this procedure does not give a conclusive choice of the "good" potential. This is in contrast with the corresponding situation for the choice of grammars, essentially because the total number of different grammars is finite.

Stronger results are available in the case of finite range potentials, which can also be used as approximate guesses to the unknown departure potential, as shown in Section 6. In this case we were able to prove that, for each chosen finite range r, both strategies lead to the identification of the original equivalence class of potentials in the limit of diverging observation time. Moreover, in the case where the observed sequence is produced by a potential that is not of finite range, both strategies will converge to a finite range solution that is a projection of the original potential on the set of the range r potentials.

Together with a criterion for the estimation of the range of a potential, the previous construction may give efficient criteria for the identification of Gibbs measures.

Finally, we show how relative entropy can also be used in identifying a grammar, adding then a different criterion to the ones discussed in ref. 5. This tool allows then, in the case of finite range potentials, to simultaneously identify the grammar and the potential.

The motivation to the investigation carried out in this article comes from the study of turbulence. In ref. 2 the authors analyse the constraints imposed on a dynamical system which is supposed to model an experimental situation with given (i.e., measured) statistical properties. We hope that the present work will help to carry further this analysis by sheding light on the properties of the statistics of time recurrences at different scales in phase space. Our confidence is based upon the existence of many results, see ref. 11 and references therein, that become available as soon as a Gibbs measure is fixed.

Besides the usual tools, in this work we made an extensive use of relative entropy obtained by a limit process of the Kullback–Leibler discrimination marginals. The use of this quantity was suggested to us by the reading of the interesting paper[12] of J. T. Lewis, C. E. Pfister, R. Russel and W. G. Sullivan. In the context of dynamical systems, the relative entropy has an interpretation which comes from a "relative ergodic result" that we obtain for this quantity.

This article is divided as follows. Section 2 gives a short list of definitions and results for the reader's convenience. In Section 3 we prove a relative ergodic result that is used in Section 4 to treat the problem of the identification of potentials and in Section 5 to discuss identification of grammars. Section 6 deals with the case of finite range potentials. We end up with a discussion that includes some perspectives for further research.

2. MEMORANDUM

The purpose of this section is to review, for the reader's convenience, the results we need in the next sections without giving proofs. Readers familiar with the subject can directly pass to Section 3.

2.1. Full Shift, Subshifts of Finite Type, and Grammars

We work with the space of all infinite sequences $\omega = i_1(\omega)\, i_2(\omega) \cdots$ where i_k, $k = 1, 2,...,$ belongs to a finite alphabet $\{1,..., N\}$, with $N \geqslant 2$. This set is $\{1,..., N\}^{\mathbb{N}}$ and we denote it by Ω. Together with the shift map: $\sigma: \Omega \to \Omega$, $\sigma\omega = i_2(\omega)\, i_3(\omega) \cdots$ $\forall j \in \mathbb{N}$, we define the full shift on N-symbols (Ω, σ).

If we specify in advance that a finite number of finite strings of consecutive symbols shall not be allowed, then we obtain a σ-invariant subprocess called a shift of finite type. There is no loss of generality if we consider forbidden words of length two by interpreting certain words as new symbols (for a precise statement of this result, we refer to ref. 13, p. 119).

In consequence, let G be a $N \times N$ matrix of zeros and ones where the $G_{i,j}$ is zero precisely when the word of length two $i_n(\omega)\, i_{n+1}(\omega) = ij$ is forbidden. We call G a grammar. So we define

$$\Omega_G := \{\omega\colon i_n(\omega) \in \{1,..., N\},\, G(i_n(\omega), i_{n+1}(\omega)) = 1,\, \forall n = 1, 2,...\}$$

$(\Omega_G, \sigma|_{\Omega_G})$ is called a shift of finite type (or a topological Markov chain). We will simply write (Ω_G, σ).

We shall assume that G is a primitive matrix, that is, there exists n_0 such that $n \geqslant n_0$ implies $G^n(i, j) > 0$, $\forall i, j = 1, 2,..., N$. Obviously, when all sequences are allowed (i.e., $G(i, j) = 1$, $\forall i, j$), then we get the full shift.

We suppose that all our measures are σ-invariant (i.e., if A is any μ-measurable set then $\mu(\sigma^{-1}A) = \mu(A)$) and normalized (i.e., $\mu(\Omega_G) = 1$). The set of such measures is denoted by $\mathscr{M}_\sigma(\Omega_G)$.

The set of all the sequences which coincide on the first n symbols is denoted by $[i_1(\omega)\, i_2(\omega) \cdots i_n(\omega)]$ or by $[i_1 i_2 \cdots i_n]$, and it is called a cylinder. Cylinders define the topology as well as the Borel sets.

2.2. Entropy

We recall two results about entropy. Our reference is ref. 14.

In our context, the measure-theoretic entropy, or simply the entropy, of a σ-invariant measure can be defined by

$$h(\mu) = \lim_{n \to \infty} \frac{H_n}{n} \tag{1}$$

where

$$H_n = - \sum_{i_1 \cdots i_n}^{N} \mu([i_1 \cdots i_n]) \operatorname{Log} \mu([i_1 \cdots i_n]) \tag{2}$$

The limit exists because $\{H_n\}_{n \geqslant 1}$ is a subadditive sequence (i.e., $H_{m+n} \leqslant H_m + H_n$ for all integers $m, n \geqslant 1$).

It is known that $\sup\{h(\mu): \mu \in \mathcal{M}_\sigma(\Omega_G)\}$ equals the topological entropy, which in the case of the full shift equals $\operatorname{Log} N$, the logarithm of the cardinal of the alphabet.

We now state the Shannon–McMillan–Breiman theorem for ergodic measures.

Theorem (Shannon–McMillan–Breiman). Let μ be an ergodic measure on (Ω_G, σ) and define

$$f_n(\omega) := -\frac{1}{n} \operatorname{Log} \mu([i_1(\omega)\, i_2(\omega) \cdots i_n(\omega)])$$

Then $\{f_n(\omega)\}_{n \geqslant 1}$ converges for almost all ω to $h(\mu)$.

The mathematical expectation of $f_n(\omega)$ is easily found by multiplying its value on each cylinder $[i_1 \cdots i_n]$ by the measure $\mu([i_1 \cdots i_n])$ of the cylinder and summing on all possible cylinders:

$$\mathcal{M} f_n(\omega) = -\frac{1}{n} \sum_{i_1 i_2 \cdots i_n}^{N} \mu([i_1 i_2 \cdots i_n]) \operatorname{Log} \mu([i_1 i_2 \cdots i_n]) = \frac{H_n}{n}$$

But $\lim_{n \to \infty}(H_n/n) = h(\mu)$ by (1), which means that the mathematical expectation of $f_n(\omega)$ approaches $h(\mu)$ as $n \to \infty$. So the theorem states that not only does the mathematical expectation of $f_n(\omega)$ but $f_n(\omega)$ itself converges for almost all ω to $h(\mu)$.

2.3. Gibbs Measures

We refer to ref. 1 or to ref. 15 for proofs, and to ref. 4 for more general results.

2.3.1. Potentials. As we shall see, a Gibbs measure is determined by a potential. The family of potentials we consider is, as usual, that of Hölder continuous functions.[2] Given $0 < \theta < 1$, we define a metric on Ω by $d_\theta(\omega, \omega') = \theta^p$ where p is the largest integer such that $\omega_i = \omega'_i$, $1 \leqslant i \leqslant p-1$ (with this metric, Ω is a complete separable metric space, i.e., a Polish space, and it is compact).

For a continuous function $\phi: \Omega \to \mathbb{R}$ and $n \geqslant 1$ define

$$\text{var}_n \phi = \sup\{ |\phi(\omega) - \phi(\omega')| : \omega_i = \omega'_i,\ 1 \leqslant i \leqslant n \}$$

It is easy to see that $|\phi(\omega) - \phi(\omega')| \leqslant C d_\theta(\omega, \omega')$ if and only if $\text{var}_n \phi \leqslant C\theta^n$, $n = 1, 2,...,$ where C is some positive constant.

Let $\Phi_\theta = \{\phi: \phi \text{ continuous, } \text{var}_n \phi \leqslant C\theta^n,\ n = 1, 2,..., \text{ for some } C > 0\}$: it is the space of Hölder continuous functions with respect to the metric d_θ.

Let $\mathscr{C}(\Omega)$ be the Banach space of real continuous functions on Ω with the sup-norm $|\cdot|_\infty$: if $\phi \in \mathscr{C}(\Omega)$, then $|\phi|_\infty := \sup\{ |\phi(\omega)| : \omega \in \Omega\}$.

We introduce the norm $|\cdot|_\theta$: $|\phi|_\theta := \sup\{(\text{var}_n \phi/\theta^n), n \geqslant 1\}$. Now Φ_θ is a Banach space with the norm $\|\cdot\|_\theta := |\cdot|_\infty + |\cdot|_\theta$.

If $0 < \theta < \theta' < 1$ then $\Phi_{\theta'} \supseteq \Phi_\theta$. This gives a "filtration" of the space of all Hölder continuous functions: $\Phi = \bigcup_{0 < \theta < 1} \Phi_\theta$.

Interesting classes of functions that lie in all of the Φ_θ, $0 < \theta < 1$, are the following. Let r be a positive integer. Then define, for $r \geqslant 1$,

$$\Phi_r = \{\phi: \Omega \to \mathbb{R},\ \phi(\omega) = \phi(\omega') \text{ if } \omega_n = \omega'_n \text{ for } 0 \leqslant n < r\}$$

Φ_r consists of locally constant functions depending on the first r symbols $i_1(\omega) \cdots i_r(\omega)$. We call $\phi \in \Phi_r$ a r-symbols potential. Notice that $\Phi_1 \subseteq \Phi_2 \subseteq \cdots$ and $\bigcup_{r=1}^{\infty} \Phi_r = \bigcap_{0 < \theta < 1} \Phi_\theta$.

We can use r-symbols potentials to uniformly approximate a general one. Indeed, assume that $\phi \in \Phi_\theta$ for some $0 < \theta < 1$, then clearly we can choose $\phi_r \in \Phi_r$ with

$$|\phi - \phi_r|_\infty \leqslant (\text{var}_r \phi/\theta^r)\, \theta^r \leqslant |\phi|_\theta\, \theta^r$$

2.3.2. Gibbs Measures. The next theorem states the existence and the unicity of Gibbs measures together with an explicit formula for the

[2] It implies that Gibbs measures coincide with equilibrium states.

measure of cylinders. Then we recall the conditions under which two potentials give the same Gibbs measure. Finally, we state a formula that relates a potential to the corresponding Gibbs measure.

Theorem. Suppose $\phi \in \Phi_\theta$. Then there is a unique σ-invariant measure, which we denote by μ_ϕ^G, for which one can find some constants P, c_1, c_2, with $0 < c_1 \leqslant 1 \leqslant c_2$ such that for all $n \geqslant 1$ and for all $\omega \in \Omega_G$:

$$c_1 \leqslant \frac{\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)])}{\exp(-nP + (S_n\phi)(\omega))} \leqslant c_2 \tag{3}$$

where $(S_n\phi)(\omega) := \sum_{k=0}^{n-1} \phi(\sigma^k \omega)$.

This double inequality holds for all $\omega \in \Omega_G$ and for all $n \geqslant 1$ and of course the constants c_1, c_2 and P depend on the potential and on the grammar. $P = P(\phi, G)$ is called the pressure of ϕ, and μ_ϕ^G the Gibbs measure associated to ϕ.

For f continuous, one defines the Perron–Frobenius–Ruelle operator:

$$(\mathscr{L}_\phi f)(\omega) = \sum_{\omega' :\, \sigma\omega' = \omega} e^{\phi(\omega')} f(\omega')$$

It can be proved that $\mu_\phi^G = b\nu$. b is the right eigenfunction (positive and Hölder continuous) associated to the simple maximal positive eigenvalue $\lambda = e^P$ of $\mathscr{L}_\phi$. ν is the corresponding left eigenfunction, which is a measure (i.e., $\int \mathscr{L}_\phi f \, d\nu = \lambda \int f \, d\nu$ for all f continuous).

A Gibbs measure is mixing, so in particular ergodic. The mixing property in the case of a grammar G is the consequence of the primitivity of G and it is equivalent to the fact that the shift σ is topologically mixing.

Definition. Two potentials $\phi, \psi \in \Phi_\theta$ are equivalent with respect to σ (we write $\phi \sim \psi$) if there exists a Hölder continuous function u and a constant $K \in \mathbb{R}$ such that

$$\psi(\omega) = \phi(\omega) - u(\omega) + u(\sigma\omega) + K \tag{4}$$

for all $\omega \in \Omega$. It is easy to check that "$\sim$" defines an equivalence relation.

Proposition:

$$\mu_\phi^G = \mu_\psi^G \qquad \text{if and only if} \quad \phi \sim \psi \tag{5}$$

This proposition implies that what can be identified is not a single potential, but only its equivalence class, since the physically observable quantity is the measure.

It is easy to check that if $\phi \sim \psi$, then $P(\phi, G) = P(\psi, G) + K$. If one is only interested in the measure, it is not a restriction to take a null pressure potential because if ϕ is given, we put $\psi = \phi - P(\phi, G)$. Then it follows that $P(\psi, G) = 0$ and $\psi \sim \phi$, hence $\mu_\psi^G = \mu_\phi^G$.

We can even define $\psi \sim \phi$ in such a way that $P(\psi, G) = 0$ and $b_\psi(\omega)$, the eigenfunction of $\mathscr{L}_\psi$ associated to the eigenvalue $e^{P(\psi, G)} = 1$, is identically one.[15] We call such at potential ψ a normalized potential. If ϕ is given, then the normalized ψ is defined by (4) with $K = -P(\phi, G)$ and $u(\omega) = -\mathrm{Log}(b_\phi(\omega))$.

Lemma (see ref. 15). If ψ is a normalized potential, then uniformly

$$\lim_{n \to \infty} \mathrm{Log} \frac{\mu_\psi^G([i_1(\omega) \cdots i_n(\omega)])}{\mu_\psi^G([i_2(\omega) \cdots i_n(\omega)])} = \psi(\omega) \tag{6}$$

Remark (1-symbol potentials and Bernoulli measures). Let us consider the full shift with $\Omega = \{0, 1\}^{\mathbb{N}}$. When ϕ is a 1-symbol potential, we have:

$$\mu([i_1]) = \frac{e^{\phi(i_1)}}{\lambda}, \qquad i = 0, 1$$

and

$$\mu([i_1 \cdots i_n]) = \mu([i_1]) \times \cdots \times \mu([i_n])$$

The measure is completly determined by:

$$\mu([1]) = \frac{1}{1 + e^{-(\phi(1) - \phi(0))}}, \qquad \mu([0]) = 1 - \mu([1]) = \frac{1}{1 + e^{(\phi(1) - \phi(0))}}$$

By putting $\mu_0 = p$, we get all the Bernoulli measures μ_p, $0 < p < 1$. We can thus represent a Bernoulli measure via a 1-symbol potential.

2.4. The Pressure and the Variational Principle

The pressure can be obtained via a variational principle:

Theorem (Variational principle). Fix a potential $\phi \in \Phi_\theta$ and a grammar G. Then the supremum $\sup_\eta\{h(\eta) + \int \phi \, d\eta\}$, taken over all σ-invariant measures, is reached only by the Gibbs measure μ_ϕ^G and equals $P(\phi, G)$:

$$P(\phi, G) = h(\mu_\phi^G) + \int \phi \, d\mu_\phi^G \tag{7}$$

This links pressure with entropy. It is important to notice that the variational principle does not make explicit reference to the dynamics.

3. A RELATIVE ERGODIC RESULT

3.1. Relative Entropy of Two Measures

In this section, following ref. 12, we define the relative entropy of two measures from the Kullback–Leibler discrepancy.[16, 17] We prove what we call a "relative ergodic result" for the relative entropy, which gives an interpretation of this quantity in the context of dynamical systems.

Definition (Kullback–Leibler discrepancy). The K. L. discrepancy of the measure λ with respect to the measure ρ is given by

$$D(\lambda \mid \rho) := \int \operatorname{Log}\left(\frac{d\lambda}{d\rho}\right) d\lambda \tag{8}$$

if λ is absolutely continuous with respect to ρ, that is if the Radon–Nikodym derivative $d\lambda/d\rho$ exists. If not, $D(\lambda \mid \rho) := +\infty$.

Notice that in general $D(\lambda \mid \rho) \neq D(\rho \mid \lambda)$ (so D is not a distance). In particular, if we have two ergodic measures we get $+\infty$ for D because they are orthogonal.

We can always define for any integer $n \geqslant 1$ the K. L. discrepancy of the corresponding n-marginals:

$$D_n(\lambda \mid \rho) := \sum_{i_1 \cdots i_n} \operatorname{Log}\left(\frac{\lambda([i_1 \cdots i_n])}{\rho([i_1 \cdots i_n])}\right) \lambda([i_1 \cdots i_n]) \tag{9}$$

We have either to suppose that ρ gives a positive measure to any cylinder of any length, or to suppose that $\lambda([i_1 \cdots i_n]) = 0$ whenever $\rho([i_1 \cdots i_n]) = 0$ because we can put "0 Log 0/0 = 0." Gibbs measures, for example, give a positive measure to any cylinder.

In the case when ρ is a Bernoulli measure and λ is any σ-invariant measure, the sequence $D_n(\lambda \mid \rho)/n$ converges. Here is the proof. Let us put $D_n := D_n(\lambda \mid \rho)$ where λ is any measure, ρ is a Bernoulli measure, i.e., $\rho([i_1 \cdots i_n]) = \rho([i_1]) \times \cdots \times \rho([i_n])$. We have

$$\begin{aligned} D_n &= -H_n(\lambda) - \sum_{i_1 \cdots i_n} \lambda([i_1 \cdots i_n]) \operatorname{Log} \rho([i_1 \cdots i_n]) \\ &= -H_n(\lambda) - n \sum_{i_1 \cdots i_n} \lambda([i_1 \cdots i_n]) \operatorname{Log} \rho([i_1]) \\ &= -H_n(\lambda) - n \sum_{i} \lambda([i]) \operatorname{Log} \rho([i]) \end{aligned}$$

For the second equality we used the σ-invariance of the measure ρ. For the third equality we used the σ-invariance of the measure λ: $\sum_{i_2 \cdots i_n} \lambda([i_1 \cdots i_n]) = \lambda([i_1])$. Hence, by (1), we obtain

$$h(\lambda \mid \rho) = \lim_{n \to \infty} \frac{D_n}{n} = -h(\lambda) - \sum_{i=1}^{N} \lambda([i]) \operatorname{Log} \rho([i]) \tag{10}$$

This motivates the following general definition:[12]

Definition (Relative entropy of two measures). The relative entropy of the measure λ with respect to ρ is given by

$$h(\lambda \mid \rho) := \limsup_{n \to \infty} \frac{1}{n} D_n(\lambda \mid \rho) \tag{11}$$

We always have $D_n(\lambda \mid \rho) \geqslant 0$ for any $n \geqslant 1$. Indeed, for any $x \geqslant 0$, $x \operatorname{Log} x \geqslant x - 1$, with equality if and only if $x = 1$. Hence for any $n \geqslant 1$, we have

$$\frac{\lambda([i_1 \cdots i_n])}{\rho([i_1 \cdots i_n])} \operatorname{Log} \frac{\lambda([i_1 \cdots i_n])}{\rho([i_1 \cdots i_n])} \geqslant \frac{\lambda([i_1 \cdots i_n])}{\rho([i_1 \cdots i_n])} - 1$$

with equality if and only if $\lambda([i_1 \cdots i_n]) = \rho([i_1 \cdots i_n])$. Multiplying this inequality by $\rho([i_1 \cdots i_n])$ and summing over $i_1,..., i_n$ yields $D_n(\lambda \mid \rho) \geqslant 0$, with equality if and only if $\lambda([i_1 \cdots i_n]) = \rho([i_1 \cdots i_n])$, $\forall i_1,..., i_n$, i.e., if and only if the marginals of the measures λ and ρ on $\{1,..., N\}^n$ are equal.

In general, $h(\lambda \mid \rho) = 0$ does not imply $\lambda = \rho$. However, we shall see that if λ is an ergodic measure and ρ a Gibbs measure, then $h(\lambda \mid \rho) = 0$ if and only if $\lambda = \rho$.

We now give two simple examples.

Example 1. Let us consider the full shift on $\Omega = \{1,..., N\}^{\mathbb{N}}$. μ is an arbitrary ergodic measure and ν the uniform product measure which can be considered as a Gibbs measure corresponding to a constant potential. Because $\nu([i]) = 1/N$, we get by (10): $h(\mu \mid \nu) = \operatorname{Log} N - h(\mu)$.

Example 2. Consider μ_p and μ_q two Bernoulli measures $(0 < p, q < 1)$ defined on the full shift with $\Omega = \{0, 1\}^{\mathbb{N}}$. We get by (10):

$$h(\mu_p \mid \mu_q) = p \operatorname{Log}\left(\frac{p}{q}\right) + (1-p) \operatorname{Log}\left(\frac{1-p}{1-q}\right)$$

3.2. A Relative Ergodic Result

Let μ_p and μ_q be two Bernoulli measures $(0 < p, q < 1)$ defined on the full shift with $\Omega = \{0, 1\}^{\mathbb{N}}$.

Let ω be a generic point of the measure μ_p, i.e., such that $\sharp\{i_k(\omega) = 0, k = 1,..., n\}/n \to p$ when $n \to \infty$, which simply means that the frequency of 0's in the sequence ω asymptotically equals p.

Thus we can write:

$$\frac{1}{n} \operatorname{Log} \mu_q([i_1(\omega) \cdots i_n(\omega)]) = \frac{\sharp\{i_k(\omega) = 0\}}{n} \operatorname{Log} q + \frac{n - \sharp\{i_k(\omega) = 0\}}{n} \operatorname{Log}(1 - q)$$

i.e.,

$$\lim_{n \to \infty} \frac{1}{n} \operatorname{Log} \mu_q([i_1(\omega) \cdots i_n(\omega)]) = p \operatorname{Log} q + (1 - p) \operatorname{Log}(1 - q), \mu_p \text{ a.e.}$$

On the other hand, we have by the Shannon–McMillan theorem that, for μ_p almost all ω:

$$\lim_{n \to \infty} \frac{1}{n} \operatorname{Log} \mu_p([i_1(\omega) \cdots i_n(\omega)]) = -h(\mu_p)$$

where $h(\mu_p) := -p \operatorname{Log} p - (1 - p) \operatorname{Log}(1 - p)$. So we get

$$\lim_{n \to \infty} \frac{1}{n} \operatorname{Log} \frac{\mu_p([i_1(\omega) \cdots i_n(\omega)])}{\mu_q([i_1(\omega) \cdots i_n(\omega)])} = h(\mu_p \mid \mu_q) \tag{12}$$

Hence, if we take a cylinder which approaches a generic point of the measure μ_p, then its measure by means of μ_q converges to zero exponentially faster than by means of μ_p. The exponential rate is precisely equal to the relative entropy $h(\mu_p \mid \mu_q) > 0$.

It is possible to generalize this idea. We replace μ_p with any ergodic measure and μ_q with any Gibbs measure in the formula (12).

Proposition. Let μ an ergodic measure and ν_ϕ^G a Gibbs measure of potential ϕ, both defined on Ω_G, for any grammar G. Then, for μ almost all ω:

$$\frac{\mu([i_1(\omega) \cdots i_n(\omega)])}{\nu_\phi^G([i_1(\omega) \cdots i_n(\omega)])} \asymp e^{nh(\mu \mid \nu_\phi^G)},$$

with

$$h(\mu \mid \nu_\phi^G) = P(\phi, G) - \int \phi \, d\mu - h(\mu) \tag{13}$$

By the notation "$\asymp$" we mean that:

$$\lim_{n \to \infty} \frac{1}{n} \operatorname{Log} \frac{\mu([i_1(\omega) \cdots i_n(\omega)])}{\nu_\phi^G([i_1(\omega) \cdots i_n(\omega)])} = h(\mu \mid \nu_\phi^G) \tag{14}$$

This proposition says that not only does the mathematical expectation of $(1/n) \operatorname{Log}(\mu([i_1(\omega) \cdots i_n(\omega)])/\nu_\phi^G([i_1(\omega) \cdots i_n(\omega)])) =: r_n(\omega)$ but $r_n(\omega)$ itself converges for μ almost all ω to $h(\mu \mid \nu_\phi^G)$.

Proof. It will be done in two steps. First, we compute $h(\mu \mid \nu_\phi^G)$, then we compute $\lim_{n \to \infty}(1/n) \operatorname{Log}(\mu([i_1(\omega) \cdots i_n(\omega)])/\nu_\phi^G([i_1(\omega) \cdots i_n(\omega)]))$.

1. For convenience, we will write simply "$\sum$" for "$\sum_{i_1 \cdots i_n}$" in our estimations. Using (3), using the facts that $\sum \mu([i_1 \cdots i_n]) = 1$, for all $n \geqslant 1$ and $H_n = -\sum \mu([i_1(\omega) \cdots i_n(\omega)]) \operatorname{Log}(\mu([i_1(\omega) \cdots i_n(\omega)]))$, we get:

$$\frac{\operatorname{Log}(c_2^{-1})}{n} \leqslant \frac{1}{n} D_n(\mu \mid \nu_\phi^G) + \frac{1}{n} H_n - P(\phi, G) + \frac{1}{n} \sum \mu([i_1(\omega) \cdots i_n(\omega)])(S_n\phi)(\omega) \leqslant \frac{\operatorname{Log}(c_1^{-1})}{n}$$

We know that $\lim_{n \to \infty}(H_n/n) = h(\mu)$ (see (1)).

Now put $g_n(\omega) := (S_n\phi)(\omega)/n$. Our goal is to prove that

$$\lim_{n \to \infty} \sum \mu([i_1(\omega) \cdots i_n(\omega)])(S_n\phi)(\omega)/n = \int \phi \, d\mu$$

We want to use the fact that for a uniformly bounded sequence of random variables, convergence almost everywhere implies convergence in mathematical expectation.[18]

Define $\tilde{g}_n(\omega) := (S_n\phi_n)(\omega)/n$ where $\phi_n \in \Phi_n$ is a n-symbols potential. We can always choose ϕ_n in such a way that $|\phi - \phi_n|_\infty \leqslant |\phi|_\theta \, \theta^n$ (see Section 2). This implies that

$$|g_n - \tilde{g}_n|_\infty = \frac{1}{n} |S_n(\phi - \phi_n)|_\infty \leqslant |\phi|_\theta \, \theta^n$$

Hence $g_n(\omega)$ and $\tilde{g}_n(\omega)$ converge almost everywhere to the same limit which, by the Birkhoff ergodic theorem, is $\lim_{n\to\infty}((S_n\phi)(\omega)/n)=\int \phi\, d\mu$, μ a.e.. Now $\sum \mu([i_1(\omega)\cdots i_n(\omega)])\,\tilde{g}_n(\omega)=:\mathscr{M}\{\tilde{g}_n(\omega)\}$ is the mathematical expectation of $\tilde{g}_n(\omega)$ which converges to $\int \phi\, d\mu$, because $\forall n$, $\forall\omega$, $|\tilde{g}_n(\omega)|\leqslant \|\phi\|_\theta$, i.e., $\tilde{g}_n$ is uniformly bounded.

It remains to write that

$$\left|\sum \mu([i_1(\omega)\cdots i_n(\omega)])\, g_n(\omega)-\sum \mu([i_1(\omega)\cdots i_n(\omega)])\,\tilde{g}_n(\omega)\right|$$

$$=\left|\sum \mu([i_1(\omega)\cdots i_n(\omega)])\,\frac{S_n}{n}(\phi-\phi_n)(\omega)\right|$$

$$\leqslant\left|\frac{S_n}{n}(\phi-\phi_n)\right|_\infty\times\sum \mu([i_1(\omega)\cdots i_n(\omega)])\leqslant|\phi|_\theta\,\theta^n$$

Hence $\lim_{n\to\infty}\mathscr{M}\{(S_n\phi)(\omega)/n\}=\lim_{n\to\infty}\mathscr{M}\{(S_n\phi_n)(\omega)/n\}=\int\phi\, d\mu$.

Finally, we get

$$h(\mu\mid\nu_\phi^G)=P(\phi,G)-\int\phi\, d\mu-h(\mu) \tag{15}$$

2. We now compute

$$\lim_{n\to\infty}\frac{1}{n}\operatorname{Log}\frac{\mu([i_1(\omega)\cdots i_n(\omega)])}{\nu_\phi^G([i_1(\omega)\cdots i_n(\omega)])}$$

By (3) we have

$$c_2^{-1}\exp(nP(\phi,G)-(S_n\phi)(\omega))\leqslant\frac{1}{\nu_\phi^G([i_1(\omega)\cdots i_n(\omega)])}$$
$$\leqslant c_1^{-1}\exp(nP(\phi,G)-S_n\phi(\omega))$$

So we get

$$\frac{\operatorname{Log}c_2^{-1}}{n}\leqslant\frac{1}{n}\operatorname{Log}\frac{\mu([i_1(\omega)\cdots i_n(\omega)])}{\nu_\phi^G([i_1(\omega)\cdots i_n(\omega)])}-P(\phi,G)$$
$$+\frac{(S_n\phi)(\omega)}{n}-\frac{\operatorname{Log}\mu([i_1(\omega)\cdots i_n(\omega)])}{n}\leqslant\frac{\operatorname{Log}c_1^{-1}}{n}$$

By the Shannon–McMillan–Breiman theorem,

$$\lim_{n\to\infty}\frac{1}{n}\operatorname{Log}\mu([i_1(\omega)\cdots i_n(\omega)])$$
$$= -h(\mu), \qquad \text{for} \quad \omega\in\tilde{\Omega} \qquad \text{such that} \quad \mu(\tilde{\Omega})=1.$$

On the other hand, by the Birkhoff ergodic theorem we have

$$\lim_{n\to\infty}\frac{1}{n}(S_n\phi)(\omega)=\int\phi\,d\mu, \qquad \text{for} \quad \omega\in\Omega_0 \qquad \text{such that} \quad \mu(\Omega_0)=1.$$

Hence we obtain, for any $\omega\in\tilde{\Omega}\cap\Omega_0(\mu(\tilde{\Omega}\cap\Omega_0)=1)$, that:

$$\lim_{n\to\infty}\frac{1}{n}\operatorname{Log}\frac{\mu([i_1(\omega)\cdots i_n(\omega)])}{\nu_\phi^G([i_1(\omega)\cdots i_n(\omega)])}=P(\phi, G)-\int\phi\,d\mu-h(\mu)=h(\mu\mid\nu_\phi^G) \quad \blacksquare$$

Remark. The pressure of a measure μ on Ω_G is defined in the following way:

$$P_\mu(\phi, G):=h(\mu)+\int\phi\,d\mu$$

So we can write the variational principle (7) under the form

$$\sup\{P_\mu(\phi, G): \mu\in\mathcal{M}_\sigma(\Omega_G)\}=P(\phi, G).$$

We can see $P_\mu(\phi, G)$ as a generalisation of the entropy $h(\mu)$ in the sense that $P_\mu(0, G)=h(\mu)$. Then, because the pressure of the null potential is the topological entropy, the variational principle generalises the variational principle for entropy because we have

$$\sup\{P_\mu(0, G): \mu\in\mathcal{M}_\sigma(\Omega_G)\}=P(0, G)=h_{top}.$$

Now, the relative entropy equals the difference between $P(\phi, G)$ and $P_\mu(\phi, G)$: $h(\mu\mid\nu_\phi^G)=P(\phi, G)-P_\mu(\phi, G)$.

Lemma. $h(\mu\mid\nu_\phi^G)=0$ if and only if $\mu=\nu_\phi^G$.

Proof. It is nothing but the variational principle (7). $\blacksquare$

3.3. Case of Two Gibbs Measures

Corollary. Let $\mu_\phi^G, \mu_{\phi'}^G$ two Gibbs measures. Then

$$h(\mu_\phi^G \mid \mu_{\phi'}^G) = P(\phi', G) - h(\mu_\phi^G) - \int \phi' \, d\mu_\phi^G$$

$$= h(\mu_{\phi'}^G) - h(\mu_\phi^G) + \int \phi' \, d\mu_{\phi'}^G - \int \phi' \, d\mu_\phi^G \tag{16}$$

Proof. It is a consequence of (15) for the first equality, of (7) for the second one. ▮

Remark. According to (5) and the last lemma, $h(\mu_\phi^G \mid \mu_{\phi'}^G) = 0$ if and only if $\phi \sim \phi'$.

Now we examine the case when we have two Gibbs measures with distinct grammars. We fix a potential $\phi \in \Phi_\theta$ and two grammars G, G' such that $G \prec G'$ (i.e., $\Omega_G \subset \Omega_{G'}$). By convention, if a cylinder is forbidden by a grammar, we write zero for the corresponding measure. Then $\mu_\phi^{G'}([i_1(\omega) \cdots i_n(\omega)]) = 0$ implies $\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)]) = 0$. For all $n \geqslant 1$, we define

$$D_n = \sum_{i_1 \cdots i_n} \mu_\phi^G([i_1 \cdots i_n]) \operatorname{Log}\left(\frac{\mu_\phi^G([i_1 \cdots i_n])}{\mu_\phi^{G'}([i_1 \cdots i_n])}\right)$$

It is well defined if we put, as it is usually done, "0 Log(0/0) = 0." Thus proposition (13) is valid:

$$\frac{\mu_\phi^G([i_1 \cdots i_n])}{\mu_\phi^{G'}([i_1 \cdots i_n])} \asymp e^{nh(\mu_\phi^G \mid \mu_\phi^{G'})}, \quad \mu_\phi^G - a.e..$$

We have obviously that

$$h(\mu_\phi^G \mid \mu_\phi^{G'}) = h(\mu_\phi^{G'}) - h(\mu_\phi^G) + \int \phi \, d\mu_\phi^{G'} - \int \phi \, d\mu_\phi^G = P(\phi, G') - P(\phi, G).$$

We immediately deduce that if $G \prec G'$ then $P(\phi, G) < P(\phi, G')$ because relative entropy is positive.

4. IDENTIFICATION OF POTENTIALS

Consider any sequence ω in Ω_G. We can ask the question if it is possible to determine the potential whose associated Gibbs measure is the best adapted to the description of the source that has produced ω. We propose two criteria for characterizing the potential ϕ we are looking for.

In practice, our knowledge of a sequence ω is limited to its n first entries. Now, if the string ω is typical for the Gibbs measure μ_ϕ^G, we expect that the weight given by μ_ϕ^G to $[i_1(\omega)\cdots i_n(\omega)]$ will be bigger than the weight given by any other Gibbs measure μ_ψ^G, if n is sufficiently large. This is of course a consequence of the relative ergodic theorem stated in Section 3.

Let us define the cyclic empirical measure T_n^ω, relative to a sequence ω which is known up to its nth entry. In the case where there is no grammar on Ω, we have

$$T_n^\omega = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{\sigma^k \mathscr{P}_n(\omega)} \tag{17}$$

Here $\mathscr{P}_n \colon \Omega \to \Omega$ is the blocking operator, defined by

$$\mathscr{P}_n(\omega) = (i_1(\omega)\, i_2(\omega)\cdots i_n(\omega)\, i_1(\omega)\, i_2(\omega)\cdots i_n(\omega)\cdots)$$

The definition of T_n^ω is such that, if $[j_1\cdots j_m]$ is any cylinder of length $m\leqslant n$, then $T_n^\omega([j_1\cdots j_m])$ is the relative frequency of the block $j_1\cdots j_m$ in the cyclic sequence $\mathscr{P}_n(\omega)$. T_n^ω is obviously ergodic.

This definition has to be slightly modified when a grammar G is present, since the block $[i_n(\omega)\, i_1(\omega)]$ could be forbidden by G. If this is the case, $\mathscr{P}_n(\omega)$ has to be defined by inserting between $i_n(\omega)$ and $i_1(\omega)$ a block of symbols $[b_1^{(n)}\cdots b_{g(n)}^{(n)}]$, chosen in a way to make $\mathscr{P}_n(\omega)$ admissible by G; of course, $n+g(n)$ replaces n in equation (17). Now, the number $g(n)$ is bounded for all n by the n_0 defined in Section 2.1, so that the influence of this modification on the measure T_n^ω will become negligible when n is large. For simplicity, the proofs where T_n^ω is involved are carried out as if it were always defined by equation (17).

If the sequence ω is typical for the Gibbs measure μ_ϕ^G, we expect T_n^ω to become closer to μ_ϕ^G than to any other Gibbs measure μ_ψ^G as n grows. Now, we know that when two measures take close values on all cylinders, their relative entropy $h(\mu_\phi^G \mid \mu_\psi^G)$ is small (according to the lemma of Section 3.2, it is zero if and only if $\mu_\phi^G = \mu_\psi^G$). So we expect the relative entropy of T_n^ω and μ_ϕ^G to become smaller than the relative entropy of T_n^ω and any other measure μ_ψ^G (with ψ not equivalent to ϕ), if n is sufficiently large.

Here are the results we have in this direction.

Proposition 1 (Maximum likelihood criterion). For any two non equivalent potentials $\phi, \psi \in \Phi_\theta$, it exists a positive integer N (which depends on ϕ, ψ, ω) such that $n \geqslant N$ implies:

$$\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)]) > \mu_\psi^G([i_1(\omega) \cdots i_n(\omega)]),$$
$$\text{for } \mu_\phi^G \text{ almost all choices of } \omega.$$

Proposition 2 (Minimum relative entropy criterion). For any two non equivalent potentials $\phi, \psi \in \Phi_\theta$, it exists a positive number N (which depends on ϕ, ψ, ω) such that $n \geqslant N$ implies:

$$h(T_n^\omega \mid \mu_\phi^G) < h(T_n^\omega \mid \mu_\psi^G), \qquad \text{for } \mu_\phi^G \text{ almost all choices of } \omega.$$

Propositions 1 and 2 essentially tell us that one is able to exclude any "wrong" potential ψ, knowing $[i_1(\omega) \cdots i_n(\omega)]$ up to a sufficiently large n. Unfortunately, this does not give a way to construct the "good" potential ϕ. Moreover, ϕ cannot be found by excluding all the other potentials ψ, since N, in Propositions 1 and 2, depends on ψ itself.

Proof of Proposition 1. It follows directly from proposition (13) in Section 3.

$$\lim_{n \to \infty} \frac{1}{n} \operatorname{Log} \frac{\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)])}{\mu_\psi^G([i_1(\omega) \cdots i_n(\omega)])} = h(\mu_\phi^G \mid \mu_\psi^G),$$
$$\text{for } \mu_\phi^G \text{ almost all choices of } \omega$$

with the observation that $h(\mu_\phi^G \mid \mu_\psi^G) > 0$ for non equivalent potentials ϕ, ψ. ∎

Proof of Proposition 2. We have

$$h(T_n^\omega \mid \mu_\phi^G) - h(T_n^\omega \mid \mu_\psi^G) = \lim_{m \to \infty} \frac{1}{m} \sum T_n^\omega([i_1 \cdots i_m])[\log \mu_\psi^G([i_1 \cdots i_m]) - \log \mu_\phi^G([i_1 \cdots i_m])] \tag{18}$$

where the sum runs over all cylinders of length m. Since this sequence converges, because of the proposition of Section 3.2, any subsequence converges to the same limit. Let us consider (18) for $m = np$, where $p = 1, 2, 3,\ldots$. Now, from the form of the measure (17) we see that there are only n cylinders of length np that have nonzero measure w.r.t. T_n^ω, namely $i_1(\omega)\, i_2(\omega) \cdots i_n(\omega)\, i_1(\omega) \cdots i_n(\omega) \cdots i_1(\omega) \cdots i_n(\omega)$, where the block $i_1(\omega)\, i_2(\omega) \cdots i_n(\omega)$ is repeated p times, and its first $(n-1)$ circular permutations. Let us call these cylinders $c_i^{(np)}(\omega)$, $i=1,\ldots,n$.

Depending on the form of $[i_1(\omega)\cdots i_n(\omega)]$, some of the $c_i^{(np)}(\omega)$ may eventually coincide. If two or more of these are equal, we will sum on only one of them, let us say the one with minimum index i. We can thus write

$$h(T_n^\omega \mid \mu_\phi^G)-h(T_n^\omega \mid \mu_\psi^G)$$
$$=\lim_{p\to\infty}\frac{1}{np}\sum_{i\in I} T_n^\omega(c_i^{(np)}(\omega))[\operatorname{Log}\mu_\psi^G(c_i^{(np)}(\omega))-\operatorname{Log}\mu_\phi^G(c_i^{(np)}(\omega))]$$

where I is included in $\{1,2,3,\ldots,n\}$.

Gibbs inequality gives

$$\operatorname{Log}\mu_\psi^G(c_i^{(np)}(\omega))\leqslant \operatorname{Log} C_\psi-npP(\psi,G)+\sum_{k=0}^{np-1}\psi(\sigma^k\omega^{(i)})$$

$$\operatorname{Log}\mu_\phi^G(c_i^{(np)}(\omega))\geqslant -\operatorname{Log} C_\phi-npP(p,G)+\sum_{k=0}^{np-1}\phi(\sigma^k\omega^{(i)})$$

where $C_\psi, C_\phi>1$, and $\omega^{(i)}$ is any sequence which belongs to $c_i^{(np)}(\omega)$. Choosing $\omega^{(i)}=\sigma^{i-1}\mathscr{P}_n(\omega)$, and using the fact that

$$\sum_{i\in I} T_n^\omega(c_i^{(np)}(\omega))=1 \tag{19}$$

we get

$$h(T_n^\omega \mid \mu_\phi^G)-h(T_n^\omega \mid \mu_\psi^G)\leqslant P(\phi,G)-P(\psi,G)+\lim_{p\to\infty}\frac{1}{np}\sum_{i=1}^{n} T_n^\omega(c_i^{(np)}(\omega))$$
$$\times\sum_{k=0}^{np-1}[\psi(\sigma^{k+i-1}\mathscr{P}_n(\omega))-\phi(\sigma^{k+i-1}\mathscr{P}_n(\omega))]. \tag{20}$$

We now need the following lemma:

Lemma. For all $\varphi\in\Phi_\theta$ and for μ_φ almost all ω

$$\lim_{p\to\infty}\frac{1}{np}\sum_{i\in I} T_n^\omega(c_i^{(np)}(\omega))\sum_{k=0}^{np-1}\varphi(\sigma^{k+i-1}\mathscr{P}_n(\omega))=\int\varphi\, dT_n^\omega, \tag{21}$$

$$\lim_{n\to\infty}\int\varphi\, dT_n^\omega=\int\varphi\, d\mu_\varphi. \tag{22}$$

Proof of (21). From definition (17) it is clear that we can write:

$$\int \psi \, dT_n^\omega = \frac{1}{n} \sum_{j=0}^{n-1} \psi(\sigma^j \mathscr{P}_n(\omega)).$$

On the other hand, using repeatedly the fact that $\sigma^{n+j}\mathscr{P}_n(\omega) = \sigma^j \mathscr{P}_n(\omega)$, $\forall j = 0, 1, 2,...,$ we get

$$\begin{aligned}
&\lim_{p\to\infty} \frac{1}{np} \sum_{i\in I} T_n^\omega(c_i^{(np)}(\omega)) \sum_{k=0}^{np-1} \psi(\sigma^{k+i-1}\mathscr{P}_n(\omega)) \\
&= \lim_{p\to\infty} \frac{1}{n} \sum_{i\in I} T_n^\omega(c_i^{(np)}(\omega)) \sum_{k=0}^{n-1} \psi(\sigma^{k+i-1}\mathscr{P}_n(\omega)) \\
&= \frac{1}{n} \sum_{i\in I} T_n^\omega(c_i^{(np)}(\omega)) \sum_{j=0}^{n-1} \psi(\sigma^j \mathscr{P}_n(\omega)) \\
&= \frac{1}{n} \sum_{j=0}^{n-1} \psi(\sigma^j \mathscr{P}_n(\omega))
\end{aligned}$$

the last equality following from (19). ∎

Proof of (22). Since $\psi \in \Phi_\theta$, we have

$$\begin{aligned}
\sup_{\omega\in\Omega} \left| \int \psi \, dT_n^\omega - \frac{1}{n} \sum_{k=0}^{n-1} \psi(\sigma^k\omega) \right| &= \sup_{\omega\in\Omega} \frac{1}{n} \left| \sum_{k=0}^{n-1} [\psi(\sigma^k \mathscr{P}_n(\omega)) - \psi(\sigma^k\omega)] \right| \\
&\leqslant \frac{1}{n} \sum_{k=0}^{n-1} \sup_{\omega\in\Omega} |\psi(\sigma^k\mathscr{P}_n(\omega)) - \psi(\sigma^k\omega)| \\
&\leqslant \frac{C}{n} \sum_{k=0}^{n-1} \theta^{n-k} \leqslant \frac{\theta}{1-\theta} \times \frac{C}{n}.
\end{aligned}$$

This implies that the sequence $\{\int \psi \, dT_n^\omega\}$ converges whenever the sequence $\{(1/n) \sum_{k=0}^{n-1} \psi(\sigma^k\omega)\}$ converges, and they have the same limit. Relation (22) now follows from the ergodic theorem applied to μ_ϕ^G. ∎

Using (20) and (21), we can write

$$h(T_n^\omega \mid \mu_\phi^G) - h(T_n^\omega \mid \mu_\psi^G) \leqslant P(\phi, G) - \int \phi \, dT_n^\omega - P(\psi, G) + \int \psi \, dT_n^\omega. \quad (23)$$

Equation (22) implies that, $\forall \varepsilon > 0$, there exists $N(\phi, \psi, \varepsilon)$ such that $n \geqslant N(\phi, \psi, \varepsilon)$ implies

$$h(T_n^\omega \mid \mu_\phi^G) - h(T_n^\omega \mid \mu_\psi^G) \leqslant P(\phi, G) - \int \phi \, d\mu_\phi^G - P(\psi, G) + \int \psi \, d\mu_\phi^G + 2\varepsilon$$

$$= -h(\mu_\phi^G \mid \mu_\psi^G) + 2\varepsilon.$$

Define for ϕ fixed $N_\psi := N(\phi, \psi, \varepsilon) = h(\mu_\phi^G \mid \mu_\psi^G)/4$. We have then

$$h(T_n^\omega \mid \mu_\phi^G) - h(T_n^\omega \mid \mu_\psi^G) < 0, \quad \forall n \geqslant N_\psi, \qquad \text{for } \mu_\phi^G \text{ almost all } \omega$$

so that the proposition is proved. ∎

5. SELECTION OF GRAMMARS REVISITED

Let us denote the set of all grammars on $\{1,..., N\}$ by $\mathscr{G}$. We can introduce a partial order in $\mathscr{G}$: $G \prec G'$ means that for all pairs (i, j) we have $G_{i,j} \leqslant G'_{i,j}$, with a strict inequality for at least one pair. It is clear that $G \prec G'$ implies $\Omega_G \subset \Omega_{G'}$.

5.1. Maximum Likelihood Procedure

In ref. 5, the following subset of $\mathscr{G}$ is defined: a potential $\phi \in \Phi_\theta$ is fixed and for all $\omega \in \Omega$ and for all $n \geqslant 1$

$$\mathscr{M}_\phi^n(\omega) = \{G: \mu_\phi^G([i_1(\omega) \cdots i_n(\omega)]) = \max_{G' \in \mathscr{G}} \mu_\phi^{G'}([i_1(\omega) \cdots i_n(\omega)])\}$$

By convention, if $[i_1(\omega) \cdots i_n(\omega)]$ is forbidden by G, we put $\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)]) = 0$. If not, we always have $\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)]) > 0$. Then we have the result:

Theorem (Ref. 5). For any potential $\phi \in \Phi_\theta$ and any grammar $G \in \mathscr{G}$, $M_\phi^n(\omega) \to \{G\}$ for μ_ϕ^G almost all ω, as n diverges.

In ref. 5, the proof is given in two lemmas. The first one eliminates all the grammars smaller than the "good" one and also those which are not comparable to G. (We recall that we only have a partial order in $\mathscr{G}$). As pointed out in ref. 5, this lemma is a direct consequence of the Birkhoff ergodic theorem for the measure μ_ϕ^G. We first recall it for the reader's convenience.

Lemma. Let G and G' two grammars such that $G' \prec G$ or such that one entry is one for G' and zero for G. Then

$$\lim_{n \to \infty} \mu_\phi^G(\{\omega: G' \in \mathscr{M}_\phi^n(\omega)\} = 0$$

It remains to eliminate grammars G' such that $G \prec G'$. This is achieved with the second lemma. Here we propose a proof that is in essence the same as in ref. 5 but that looks shorter thanks to the result (13) on relative entropy.

Lemma. For any potential $\phi \in \Phi_\theta$, there exists an integer $N_0 = N_0(\omega)$, which does not depend on G because of the finitness of $\mathscr{G}$, such that for any pair of grammars G, G' with $G \prec G'$ and for all $n \geqslant N_0$, we have for μ_ϕ^G almost all ω:

$$\mu_\phi^{G'}([i_1(\omega) \cdots i_n(\omega)]) < \mu_\phi^G([i_1(\omega) \cdots i_n(\omega)])$$

Proof. Let us fix a potential $\phi \in \Phi_\theta$. If $G \prec G'$ then we can apply proposition (13):

$$\frac{\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)])}{\mu_\phi^{G'}([i_1(\omega) \cdots i_n(\omega)])} \asymp e^{nh(\mu_\phi^G \mid \mu_\phi^{G'})} \qquad \text{for } \mu_\phi^G \text{ almost all } \omega$$

with $h(\mu_\phi^G \mid \mu_\phi^{G'}) = P(\phi, G') - P(\phi, G) > 0$. This means that given any $\varepsilon > 0$ and any ω in a set of measure one w.r.t. μ_ϕ^G, there exists an integer $N = N(\varepsilon, G, \omega)$ such that for all $n \geqslant N$ we have

$$\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)]) \geqslant \mu_\phi^{G'}([i_1(\omega) \cdots i_n(\omega)])\, e^{n(h(\mu_\phi^G \mid \mu_\phi^{G'}) - \varepsilon)}$$

Choosing an ε_0 smaller than $h(\mu_\phi^G \mid \mu_\phi^{G'})$, for $N_0 = N(\varepsilon_0, G, \omega)$ we get:

$$\mu_\phi^G([i_1(\omega) \cdots i_n(\omega)]) > \mu_\phi^{G'}([i_1(\omega) \cdots i_n(\omega)]) \quad \blacksquare$$

5.2. Minimum Relative Entropy Procedure

In ref. 5, the authors define a second procedure of identification based on the minimisation of entropy, that gives the same answer as before for potentials of sufficiently small norm. We now define another procedure, the minimum relative entropy procedure, which identifies the good grammar for any potential.

Given a sequence ω and a potential $\phi \in \Phi_\theta$, we define for all $n \geqslant 1$ the following set (the minimum relative entropy set):

$$\mathscr{E}_\phi^n(\omega) = \{G \in \mathscr{G}: [i_1(\omega) \cdots i_n(\omega)] \subset \Omega_G \text{ and } h(T_n^\omega \mid \mu_\phi^G) \text{ minimum}\}$$

T_n^ω is the cyclic empirical measure based on ω defined in Section 4 for the case where a grammar is present.

Theorem. For any potential ϕ and any grammar G, the sets $\mathscr{E}_\phi^n(\omega)$ converge to $\{G\}$ for μ_ϕ^G almost all ω as n goes to infinity.

Proof. Let G' be such that $G \prec G'$ and ω a given sequence. Using the proposition (13) of Section 3, we get:

$$h(T_n^\omega \mid \mu_\phi^G) - h(T_n^\omega \mid \mu_\phi^{G'}) = P(\phi, G) - \int_{\Omega_G} \phi \, dT_n^\omega - P(\phi, G') + \int_{\Omega_{G'}} \phi \, dT_n^\omega$$

$$= P(\phi, G) - P(\phi, G') + \int_{\Omega_{G'}/\Omega_G} \phi \, dT_n^\omega$$

On the one hand, we know that $G \prec G'$ implies $P(\phi, G) < P(\phi, G')$ for any potential (see Section 3.3). On the other hand, the facts that $\mu_\phi^G(\Omega_{G'}/\Omega_G) = 0$, and that T_n^ω converges weakly, for μ_ϕ^G almost all ω (see equation (22) of Section 4), imply that:

$\forall \varepsilon > 0$, there exists $N = N(\omega, \varepsilon)$, which does not depend on G because of the finitness of $\mathscr{G}$, such that $n \geqslant N$ implies

$$\int_{\Omega_{G'}/\Omega_G} \phi \, dT_n^\omega < \varepsilon, \qquad \text{for } \mu_\phi^G \text{ almost all } \omega$$

Now let $\varepsilon_0 = (P(\phi, G') - P(\phi, G))/2$. Then there exists $N_0 = N_0(\omega) := N(\omega, \varepsilon_0)$ such that for any $n \geqslant N_0(\omega)$,

$$h(T_n^\omega \mid \mu_\phi^G) - h(T_n^\omega \mid \mu_\phi^{G'}) < 0$$

The theorem is proved. ∎

6. FINITE-RANGE POTENTIALS

The interest of considering finite range potentials lies in the fact that, as it has been mentioned in Section 2, they can be used to uniformely approximate any Holder continuous potential ϕ^*, while the corresponding measures converge in the vague topology to μ_{ϕ^*}.[4]

We will see in this section that, on the one hand, we can give criteria to test if an unknown potential is a finite range one; on the other hand, we can propose two procedures to find what is the best approximation of range r to an unknown potential.

6.1. A Criterion for Determining the Range

An important property of r-symbol potentials, which can be derived using the Perron–Frobenius–Ruelle operator, is that the associated Gibbs measure μ satisfies the equation

$$\mu([i_1,..., i_n]) = \frac{\mu([i_1,..., i_r]) \times \mu([i_2,..., i_{r+1}]) \times \cdots \times \mu([i_{n-r+1},..., i_n])}{\mu([i_2,..., i_r]) \times \mu([i_3,..., i_{r+1}]) \times \cdots \times \mu([i_{n-r+1},..., i_{n-1}])} \tag{24}$$

for all values of $i_1,..., i_n$, $n \geqslant r$.

Equation (24) allows to derive a closed expression for the entropy $h(\mu)$ in terms of the measures of cylinders of finite length. Substituting (24) in (2), and using the relations which express the σ-invariance of the measure:

$$\sum_{i_1} \mu([i_1,..., i_n]) = \mu([i_2,..., i_n]), \qquad \sum_{i_n} \mu([i_1,..., i_n]) = \mu([i_1,..., i_{n-1}])$$

we get

$$h(\mu) = -\sum_{i_1,..., i_k} \mu([i_1,..., i_k)] \operatorname{Log} \frac{\mu([i_1,..., i_k]}{\mu([i_1,..., i_{k-1}]} = H_k - H_{k-1} \tag{25}$$

which is valid for all $k \geqslant r$, if $r > 1$. In the case $r = 1$, we have $h(\mu) = H_1 = H_k - H_{k-1}$ for all $k > 1$.

Now, imagine to be given a sequence ω produced by the dynamical system under study, which we suppose to be described by an unknown Gibbs measure. What we just discussed shows that there are ways to determine if the associated potential is a finite range one.

A possible criterion is based on equation (25). In fact, the knowledge of ω up to its nth entry allows to determine the empirical measures T_n^ω of cylinders of any length, and also to calculate

$$H_k^T = -\sum_{i_1,..., i_k} T_n^\omega([i_1,..., i_k]) \operatorname{Log} T_n^\omega([i_1,..., i_k])$$

for any k. If there is a r for which the quantity $H_k^T - H_{k-1}^T$ is close to be constant in k for $r \leqslant k \ll n$, then we can assume that the potential we are looking for has range r.

6.2. Identification of Finite-Range Potentials

The question we will now address is that of identification of potentials, in the case where we restrict our search to potentials of range r. We will obtain more conclusive results than those of Section 4. In fact, for simplicity, we shall treat the problem of simultaneously identifying the potential and the grammar.

Let us consider the case of the alphabet $\{1,..., N\}$ where no grammar is present. A Gibbs measure μ of range r is completely determined by $(N-1)\, N^{r-1}$ independent parameters. A possible choice for these is

$$p_{jK} = e^{\psi(jk_1 \cdots k_{r-1})} \tag{26}$$

where $j = 1,..., N-1$, $K = (k_1 \cdots k_{r-1})$ follows the lexicografic order of $k_1,..., k_{r-1} = 1,..., N$, and ψ is a normalized potential associated to μ. Since a normalized potential ψ satisfies the equation

$$\sum_{j=1}^{N} e^{\psi(jk_1 \cdots k_n)} = 1, \qquad \forall k_1,..., k_n = 1,..., N, \quad \forall n \tag{27}$$

ψ cannot be positive. So, the parameters $p = \{p_{jK}\}$ take values in the open set $(0, 1)^{(N-1)\, N^{r-1}}$.

Let us extend the set of possible values of the parameters $p = \{p_{jK}\}$ to the compact set $\mathscr{P} = [0, 1]^{(N-1)\, N^{r-1}}$. This amounts to adding measures for which one or more cylinders of length r are forbidden (i.e., have measure zero). If the range r is at least two, between these there are all Gibbs measures μ_ϕ^G with ϕ Hölder continuous of range r, and G any grammar of the type defined in Section 2.

For any sequence ω, for any $n \geq 1$, let us now define the Maximum Likelihood set

$$\mathscr{M}_n^r(\omega) = \{ p \in \mathscr{P} : \mu_p([i_1(\omega) \cdots i_n(\omega)]) = \max_{q \in \mathscr{P}} \{\mu_q([i_1(\omega) \cdots i_n(\omega)])\}\}$$

$\mathscr{P}$ is compact, $\mu_p([i_1(\omega) \cdots i_n(\omega)])$ is a continuous functions of the parameters p for any fixed sequence ω. This implies that the maximum is attained at least for some $p \in \mathscr{P}$.

Let us also define the Minimum Relative Entropy set

$$\mathscr{E}_n^r(\omega) = \{ p \in \mathscr{P} : h(T_n^\omega \mid \mu_p) = \min_{q \in \mathscr{P}} \{h(T_n^\omega \mid \mu_q)\}\}$$

where T_n^ω is defined according to (17). $\mathscr{P}$ is compact, $h(T_n^\omega \mid \mu_p)$ is a lower semi-continuous functions of the parameters p for any fixed sequence ω, so that the minimum is attained at least for some $p \in \mathscr{P}$.

Let us now suppose that the sequence ω is produced by a dynamical system described by an unknown Gibbs measure $\mu_{\phi^*}^{G^*}$, the one we want to identify. We have:

Proposition 1. In the limit $n \to \infty$, the set $\mathscr{M}_n^r(\omega)$ converges to a unique p_∞, for $\mu_{\phi^*}^{G^*}$ almost all ω. Moreover, the corresponding measure μ_{p_∞} equals $\mu_{\phi^*}^{G^*}$ if the potential ϕ^* is of range r.

Proposition 2. In the limit $n \to \infty$, the set $\mathscr{E}_n^r(\omega)$ converges to a unique p_∞, the same as in Proposition 1, for μ_{ϕ^*} almost all ω. The corresponding measure μ_{p_∞} equals $\mu_{\phi^*}^{G^*}$ if the potential ϕ^* is of range r.

Remark. If ϕ^* is not of range r, μ_{p_∞} represents the best approximation to $\mu_{\phi^*}^{G^*}$ in the sense of both the Maximum Likelihood and Minimum Relative Entropy criteria. In fact, μ_{p_∞} and $\mu_{\phi^*}^{G^*}$ coincide on all cylinders of length smaller or equal to r.

Proof of Proposition 1. Let us define

$$z_n(p) = \frac{1}{n} \operatorname{Log} \mu_p(\omega[n]).$$

Using the Perron–Frobenius–Ruelle equation it is easy to show that, in the case of range r, the measure of any cylinder of length $n \geqslant r$ can be expressed as

$$\mu([i_1 \cdots i_n]) = e^{\psi(i_1 \cdots i_r)} \times e^{\psi(i_2 \cdots i_{r+1})} \times \cdots \times e^{\psi(i_{n-r+1} \cdots i_n)} \times \mu([i_{n-r+2} \cdots i_n]). \tag{28}$$

Using (28) together with equations (26) and (27) we can write

$$z_n(p) = f_n(p) + g_n(p)$$

with

$$f_n(p) = \frac{1}{n} \operatorname{Log} \mu_p([i_{n-r+2}(\omega) \cdots i_n(\omega)])$$

$$g_n(p) = \sum_{K=1}^{N^{r-1}} \left[\sum_{j=1}^{N-1} \frac{\alpha_{jK}^{\omega}}{n} \operatorname{Log}(p_{jK}) + \frac{\beta_K^{\omega}}{n} \operatorname{Log}\left(1 - \sum_{j=1}^{N-1} p_{jK}\right)\right]$$

where α_{jK}^{ω} is the number of subsequences $\{jK\}$ appearing in $[i_1(\omega) \cdots i_n(\omega)]$, and β_K^{ω} is the number of subsequences $\{NK\}$ appearing in $[i_1(\omega) \cdots i_n(\omega)]$.

For all ω in a set Ω_G^* of measure one w.r.t. $\mu_{\phi^*}^{G^*}$, such that all $\omega \in \Omega_G^*$ do not contain transitions forbidden by G^*, we have the following facts:

(1) the numbers $\alpha_{jK}^{\omega}, \beta_K^{\omega}$ converge to $\mu_{\phi^*}^{G^*}([jK]), \mu_{\phi^*}^{G^*}([NK])$ for $n \to \infty$;

(2) the function

$$g_\infty(p) = \lim_{n \to \infty} g_n(p)$$
$$= \sum_{K=1}^{N^{r-1}} \left[\sum_{j=1}^{N-1} \mu_{\phi^*}^{G^*}([jK]) \operatorname{Log}(p_{jK}) + \mu_{\phi^*}^{G^*}([Nk]) \operatorname{Log}\left(1 - \sum_{j=1}^{N-1} p_{jK}\right)\right]$$

has its maximum on $\mathscr{P}$ at the point p^*, defined by

$$\begin{aligned} p_{jK}^* &= \frac{\mu_{\phi^*}^{G^*}([jK])}{\mu_{\phi^*}^{G^*}([K])} && \text{if } \mu_{\phi^*}^{G^*}([jK]) \neq 0 \\ p_{jK}^* &= 0 && \text{if } \mu_{\phi^*}^{G^*}([jK]) = 0 \end{aligned} \tag{29}$$

(3) $f_n(p)$ converges uniformely to the function zero for $n \to \infty$, in a closed ball centered at p^*;

(4) $g_n(p)$ converges uniformely to $g_\infty(p)$ in a closed ball centered at p^*.

Notice that (1) to (4) are true independently of the fact that p^* is in the interior of $\mathscr{P}$ (if G^* is the full Markov matrix) or on the border of $\mathscr{P}$.

Now, let us consider a sequence $\{p_n\}_n$ of points in which $z_n(p)$ has a maximum, which converges to a point p_∞. Points (1)–(4) imply that $g_\infty(p)$ has a maximum at p_∞, which must coincide with p^*. We have thus proved that the set $\mathscr{M}_n^r(\omega)$ converges to the unique point p_∞ defined by (29). If the potential ϕ^* is of range r, Eq. (6) implies that the measure μ_{p^∞} coincides with $\mu_{\phi^*}^{G^*}$.

Proof of Proposition 2. The measure T_n^ω is ergodic. Using the proposition (13), it can be shown that, if μ_p is the measure corresponding to a normalized potential ψ of range r:

$$h(T_n^\omega \mid \mu_p) = -h(T_n^\omega) - \sum_{i_1 \cdots i_r} \psi(i_1 \cdots i_r)\, T_n^\omega([i_1 \cdots i_r])$$
$$= \sum_{K=1}^{N^{r-1}} \left[\sum_{j=1}^{N-1} T_n^\omega([jK]) \operatorname{Log}(p_{jK}) + T_n^\omega([NK]) \operatorname{Log}\left(1 - \sum_{j=1}^{N-1} p_{jK}\right)\right]$$

The function T_n^ω has its maximum on $\mathscr{P}$ at the point $p^{(n)}$, defined by

$$p_{jK}^{(n)} = \frac{T_n^\omega([jK])}{T_n^\omega([K])} \quad \text{if} \quad T_n^\omega([jK]) \neq 0$$
$$p_{jK}^{(n)} = 0 \quad \text{if} \quad T_n^\omega([jK]) = 0 \tag{30}$$

so that

$$\mathscr{E}_n^r(\omega) = \{ p_{jK}^{(n)} \}, \quad \forall n$$

Now, the numbers $T_n^\omega([jK])$, $T_n^\omega([NK])$ converge to $\mu_{\phi*}^{G*}([jK])$, $\mu_{\phi*}^{G*}([NK])$ for $n \to \infty$, for $\mu_{\phi*}^{G*}$ almost all ω. This implies that $\mathscr{E}_n^r(\omega)$ converges to the unique point p_∞:

$$p_\infty = \frac{\mu_{\phi*}^{G*}([jK])}{\mu_{\phi*}^{G*}([K])}$$

the same as in equation (29). The sets $\mathscr{E}_n^r(\omega)$ and $\mathscr{M}_n^r(\omega)$ may not coincide at finite n but they do converge to the same unique point for $n \to \infty$.

Remark. According to the Minimum Relative Entropy criterion, the best choice of the measure at finite n is given by equation (30). Notice that this coincides with the choice of a potential of range r that one would make by directly applying equation (6), approximating the unknown asymptotic measure by the empirical one.

7. CONCLUSIONS AND PERSPECTIVES

Whether or not Gibbs measures are relevant for a dynamical modeling of turbulence remains an open question. However, since these ergodic measures have a transparent structure and very much is known about their properties, it is tempting to consider possible criteria for the identification of potentials (or Gibbs measures) from experimental data. The present work is an attempt in this direction.

The main tool used in our work is the relative entropy of an ergodic measure with respect to a Gibbs measure, a quantity related to the Kullback–Leibler discrepancy. In this way, in Section 3, we prove a relative ergodic theorem. Since this result may be read as an exponential splitting of Poincaré return times for orbits that are typical for different potentials, it is natural to use it for the identification of Gibbs measures. This is carried out in Section 4 where we can see that, for each finite time observation, "wrong" potentials are ruled out, even if the strategy may not converge for asymptotic times.

We also show how relative entropy can also be useful for the identification of grammars, adding a new strategy to the ones proposed in ref. 5.

Then we describe an alternative way to treat the problem. Since finite range potentials approach any potential exponentially fast as their range increases, and the corresponding measures also converge in the vague topology, it is worthwhile to solve the problem inside the set of finite range potentials for a given range and then eventually increase it to improve the estimation. In Section 4 we show how this can be done by two possible procedures. The first, already employed in ref. 5, uses a maximum likelihood principle: at each stage of the observation we choose, inside the potentials of range r, those who maximise the measure of the observed sequence. A second criterion uses the relative entropy of empirical measures with respect to Gibbs measures as a tool for identification.

By giving a way of estimating the range of a potential from observation, we have tried to fill the gap between Section 4 and Section 6.

As stated, it is not clear for the moment what shall be the outcome of the identification strategies when applied to experimental sequences in turbulence, an issue we are presently dealing with.

Another related topic, see ref. 19, is the study of possible interesting subsets of sequences, still of full measure, for which the convergence of the strategies of identification may be faster and relevant for turbulence.

We hope that our work will increase the present interest in linking statistics and dynamics in turbulence.

ACKNOWLEDGMENTS

The authors sincerely thank the CIC, A. C. in Cuernavaca and in particular G. Martinez-Mekler and Th. Seligman for the nice and stimulating atmosphere during the period when this paper was written. They also gratefully acknowledge support from the Nonlinear Dynamics Academic Year held in Cuernavaca, Mexico. Finally, we thank one of the referees for having drawn our attention on ref. 9.

REFERENCES

1. R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Mathematics 470, Springer-Verlag.
2. E. Ugalde and R. Lima, On a discrete dynamical model for local turbulence, *Physica D* **95**:144–157 (1996).
3. E. Ugalde, Self-similarity and finite-time intermittent effects in turbulent sequences, *J. Phys. A: Math. Gen.* **29**:4425–4443 (1996).

4. D. Ruelle, *Thermodynamic Formalism*, Encyclopedia of Mathematics and Its Applications 5 (Addison-Wesley, 1978),
5. P. Collet, A. Galves, and A, Lopes, Maximum likelihood and minimum entropy identification of grammars, *Random and Computational Dynamics* **3**:241–250 (1995).
6. P. Billingsley, *Statistical Inferences for Markov Chains* (Univ. of Chicago Press, 1961).
7. M. Katz, On some Criteria for Estimating the order of a Markov Chain, *Technometrics* **23**:243 (1982).
8. Comets *et al.*, Parameter estimation for Gibbs distributions from partially observed data, *Ann. Appl. Prob.* **2**(1):142–170 (1992).
9. A. Raftery and S. Tavare, Estimation and modeling repeated patterns in high order Markov chains with the mixture transition distribution model, *J. R. Stat. Soc., Ser. C* **43**(1):179–199 (1994).
10. D. S. Ornstein and B. Weiss, How sampling reveals a process, *Annals of Probability* **18**:905–930 (1990).
11. L. Barreira, Y. Pesin, and J. Schmeling, On a general concept of multifractality: Multifractal spectra for dimensions, entropies, and Lyapunov exponents. Multifractal rigidity, *Chaos* **7**:27–38 (1997).
12. J. T. Lewis, C.-E. Pfister, R. Russel, and W. G. Sullivan, Reconstruction sequences and equipartition pleasures: An examination of asymptotic equipartition property, to appear in *IEEE Transactions on Information Theory*.
13. M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic Theory on Compact Spaces*, Lecture Notes in Mathematics 527, Springer-Verlag.
14. M. Smorodinsky, *Ergodic Theory, Entropy*, Lecture Notes in Mathematics 214, 1971.
15. W. Parry and M. Pollicott, Zeta functions and the periodic orbit structure of hyperbolic dynamics, *Astérisque* **187-188**, SMF, 1990.
16. S. Kullback, *Information Theory and Statistics* (Wiley, 1959).
17. S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. Math. Statist.* **22**:79–86, 1951.
18. P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, third edition, 1995.
19. E. Ugalde, in preparation.

822/90/3-4-14