ON THE DIMENSION OF THE PREDICTIVE PROCESS OF A MEMORYLESS CHANNEL

Artur Lopes and Sílvia Lopes

Instituto de Matemática
Universidade Federal do Rio Grande do Sul
Porto Alegre – RS – 91540-000 Brasil

ABSTRACT

We analyze metrical properties of the unique stationary law for the one-step predictor of a finite state Markov Chain from noisy observations. In Piccioni (1990), the topological aspect of this problem was analyzed. Our work is a natural follow-up of this paper. We will be concerned with the case where the stationary law has support in a totally disconnected and perfect set. In this case the predictor keeps an infinite memory of the past observations. The closure of the support of this stationary law is called *the attractor* S (Elton and Piccioni (1992)). We present a lower bound for the dimension of S. This lower bound will be also an upper bound for the exponent scale of the law. As a consequence of our results, we partially answer a question raised by Piccioni (1990), in a case (b=c, see notation in Section 4) where the closure of the invariant measure's support is an interval, showing that the stationary law is singular with respect to the Lebesgue measure.

1. INTRODUCTION

First, we will recall the main definitions and results of the paper Elton and Piccioni (1992), where they analyze a class of Markov processes which arises in some problems of recursive estimation of Markov Chains.

Let $\{X_n\}_{n\geq 0}$ be an irreducible, aperiodic d-state Markov chain and $\{Y_n\}_{n\geq 0}$ be an observation of the process coming from a noisy memoryless channel with d possible outputs. The predictor

$$S_n(i) = P\{X_n = i| \; Y_j, j = 1, 2, \cdots, n-1\}, \quad i = 1, 2, \cdots, d$$

is a Markov process on the unit d-dimensional simplex $\sum_d$. This is so, because it is representable as $S_{n+1} = F_{Y_n}(S_n)$, with Y_n conditionally independent of $S_{n-1}, S_{n-2}, \cdots$, given S_n, where $F_j(\cdot)$, $j \in \{1, 2, \cdots, d\}$, are maps defined by Elton and Piccioni (1992).

The simplex $\sum_d$ is by definition the subset of $\mathbf{R}^d$, given by

$$s_1 + s_2 + \cdots + s_d = 1, \quad s_1, s_2, \cdots, s_d \geq 0.$$

Under quite general conditions, Elton and Piccioni (1992) show that

$$\text{diameter}(F_{i_1} \circ F_{i_2} \circ \cdots \circ F_{i_n})(\sum_d) \to 0,$$

exponentially fast, uniformly in $\{i_n\}$, as n goes to infinity. ¿From this result follows the uniqueness of the stationary law μ for $\{S_n\}_{n\geq 0}$.

The closure of the support of this stationary law will be denoted by S.

They also characterize when S is totally disconnected or not. In order to do that, they consider the Iterated Function System associated to the process $\{S_n\}_{n\geq 0}$.

We refer the reader to Elton and Piccioni (1992) for the relevant considerations about filtering problems and signal to noise ratio properties. The process $\{Y_n\}_{n\geq 0}$ is called *memoryless channel* in the language of communication theory.

We will consider here only the binary case $d = 2$ and therefore, we will use the notation of Piccioni (1990).

The paper of Elton and Piccioni (1992) extends results presented for the case $d = 2$ in Piccioni (1990). We will be concerned in analyzing the case considered by Piccioni (1990).

Given a probability θ, *the exponent scale of* θ, denoted by α, is the value such that

$$\theta(B(x, r)) \approx r^\alpha$$

for θ-almost every point $x \in X$ and for small r. As usual, $B(x, r)$ denotes the ball of center x and radius r. The precise mathematical meaning of the notation $\approx$ appears in Mañé (1990). Such value α always exists for an invariant probability (see Mañé (1990)).

We will present an upper bound for the exponent scale of μ and a lower bound for the Hausdorff dimension of S.

The main result of the present paper, Theorem 3.1, will be proved on Section 4.

In Section 2 we will recover the main results of Piccioni (1990) and Elton and Piccioni (1992) and in Section 3 we will state the main properties of thermodynamic formalism that will be needed in the sequel.

We refer the reader to Billingsley (1965) and Falconer (1990) for the definition of Hausdorff dimension of a set.

We would like to thank S. Carmona and I. Tanaka for helpful discussions about the prediction in Markov Chains.

2. THE PREDICTIVE PROCESS

Now we will introduce the precise definitions presented in Elton and Piccioni (1992).

Let $\{X_n\}_{n\geq 0}$ be a Markov chain on the state space $I_d = \{1, 2, \cdots, d\}$ with transition probability matrix $P = (p_{i,j})_{i,j\in I_d}$. The matrix P is supposed to be primitive, that is, there exists some integer N such that P^N has all positive entries.

The process $\{X_n\}_{n\geq 0}$ represents the signal and the process $\{Y_n\}_{n\geq 0}$ the observations subject to some kind of noise. The probability of an observation $i \in I_d$ given that at the same time the signal is $j \in I_d$, is ϵ_{ij}, independent of all other signal values at different times. We assume that ϵ_{ij} is bounded away from zero and one, for all $i, j \in I_d$.

The next proposition is proved in Elton and Piccioni (1992).

Proposition 2.1: *The process* $\{S_n\}_{n\geq 0}$ *is a homogeneous Markov Process on* $\sum_d$ *with transition kernel*

$$P(s, dt) = \sum_{i=1}^{d} \pi_s(i)\delta_{F_i(s)}(dt),$$

where $s = (s_1, s_2, \cdots, s_d) \in \sum_d \subset \mathbf{R}^d$ *and* $F_i : \sum_d \to \sum_d$*,* $i \in I_d$*, is defined by*

$$F_i(s)(h) = \frac{\sum_{j=1}^d p_{jh}\epsilon_{ji}s_j}{\sum_{j=1}^d \epsilon_{ji}s_j}, \qquad h \in I_d,$$

and

$$\pi_i(s) = \sum_{j=1}^d \epsilon_{ji}s_j. \tag{2.1}$$

In Elton and Piccioni (1992) it is shown that, under these hypotheses, there exists a unique stationary law μ, which is attractive, that is, $\{S_n\}_{n\geq 0}$ converges weakly to μ irrespectively of its entrancy law.

In Elton and Piccioni (1992) it is solved the topological question of characterizing when the set S, the closure of the support of μ, is a totally disconnected and perfect set.

3. THERMODYNAMIC FORMALISM

Recently, the interplay between probability and thermodynamic formalism has been explored (see Lalley (1991) and Lopes (1990)).

We will present a method based in thermodynamic formalism techniques (see Parry and Pollicott (1990) and Ruelle (1989)), that allows one to obtain estimates for the Hausdorff dimension of S and for the exponent scale of the measure μ. We will be concerned here only with the case when the set is totally disconnected and perfect. The reason for this assumption is that, in this case, there exists a map T such that, the functions F_i, defined in (2.1) for $i \in I_d$, are the inverse branches of T defined on $\sum_d$. In fact, the map T will be defined in a piece of $\sum_d$ where the non-wandering set of the Iterated Function System is contained. To be more rigorous, we should extend the map T, in any manner, to all $\sum_d$ and proceed as we will do in Section 4.

We shall state now the main properties of the thermodynamic formalism that will be considered in the sequel.

Given a map T defined on a compact metric space X, the set of invariant probabilities is denoted by $M(T)$.

For a continuous function ϕ defined for $x \in X$ and taking real values,

$$P(\phi) = \sup_{\theta \in M(T)} \{h(\theta) + \int \phi(x)d\theta(x)\} \tag{3.1}$$

is called *the topological pressure of* ϕ. As usual, $h(\theta)$ denotes *the entropy of* θ (see Mañé (1987)).

If T is expansive and ϕ is Holder, there exists just one measure ν_ϕ that attains the supremum value defined in (3.1). This measure is called *the equilibrium measure for* ϕ.

Given two functions ϕ and ψ, we say that they have *the homology property* if there exist a function η and $w \in \mathbf{R}$ such that $\phi = \psi + \eta \circ T - \eta + w$. In this case, it is easy to see that a measure ν is of equilibrium for ϕ, if and only if, ν is of equilibrium for ψ. This property will be crucial for our reasoning in Section 4.

Suppose now X is a subset of $\mathbf{R}^m$ and the expansive map T is differentiable. Then the function

$$s(t) = P(-t \log|\det DT|) \tag{3.2}$$

is differentiable, convex and for a unique value t_0, we obtain $s(t_0) = 0$, where DT denotes the derivative matrix of T. In the case $m = 1$, the value $t_0 = HD(S)$ is *the Hausdorff dimension* of the non-wandering set S of T. Furthermore, for any t, the tangent line defined by the line with slope $s'(t)$ through $s(t)$, intersects the t-axis in a unique value α_t.

¿From thermodynamic formalism theory (see, for instance, Ruelle (1989) and Lopes (1989)), it follows that α_t defined above is the exponent scale of the equilibrium state $\nu_{-t \log|\det DT|}$ (see Figure 2).

Now we can state more precisely the results presented here. Assuming the totally disconnected and perfect case one can obtain a map T such that the inverse branches are the functions $F_i, i \in I_d$, defined in (2.1). From previous estimates of Piccioni (1990) and Elton and Piccioni (1992), it follows that the map T is expansive. We will show that the stationary probability μ of the process $\{S_n\}_{n\geq 0}$ is an equilibrium state (for the topological pressure) for $\phi = \log V$, where V is defined in Section 4 by expression (4.2), (in fact μ is a g - measure; see Parry and Pollicott (1991)).

The main point in the present paper is the special relationship between $\log V$ and $\log|T'|$. Without this property, any estimate of the Hausdorff dimension of S or of the exponent scale of μ would be impossible to obtain.

The parameters a and b in Theorem 3.1 below are defined in Piccioni (1990), and they describe the probability of occurrence of noise in the model. We refer the reader to Piccioni (1990) for the precise meaning of a and b.

Theorem 3.1: *In the binary case ($d = 2$), an upper bound for the exponent scale α of the invariant measure μ and a lower bound for the Hausdorff dimension of the attractor S are given by*

$$\alpha < -\frac{\log 2}{\log a(1-b^2)} < HD(S).$$

4. THE BINARY CASE

Given the constants a and b, $0 < a, b < 1$, let c be the solution of

$$c = a\frac{c+b}{1+b}.$$

We will analyze here just the case where $b \geq c$.

Consider the mapping $T : [-c, c] \to [-c, c]$ defined by

$$T(y) = g_1(y) = \frac{y-ab}{a-by}, \quad \text{if } y \geq 0 \quad \text{and} \quad T(y) = g_2(y) = \frac{y+ab}{a+by}, \quad \text{if } y < 0.$$

One observes that the inverse branches of T are given by

$$f_+(x) = a\frac{x+b}{1+bx} \quad \text{and} \quad f_-(x) = -a\frac{b-x}{1-bx}.$$

In part (a) of Figure 1 we show the graph of f_+ and f_- in the case $b > c$. Part (b) of Figure 1 shows the case $b = c$ and finally part (c) of the same figure shows the case $b < c$. In the last case it is not possible to find T such that f_+ and f_- are inverse branches of T.

This is the reason why we consider only the cases $b \geq c$. All these figures show the graph of f_+ and f_- on the interval $[-c, c]$. We plot the diagonal line with dotted line.

Lemma 3 in Piccioni (1990) (see page 327) claims that the mapping T is expansive and Theorem 1 (see page 323) shows the existence of a stationary probability ν. As before, S denotes the closure of the support of ν.

The fixed point of f_+ is c. There is no need to consider f_+ and f_- defined on $[-1, 1]$. It is enough to consider these functions defined on $[-c, c]$ since the dynamics of the iterated systems f_+ and f_- are concentrated in $[-c, c]$. Piccioni (1990) shows that for $b > c$ the set S is a Cantor set and for $b \leq c$, S is an interval. We will consider here the case $c \leq b$ where f_+ and f_- are the inverse branches of the map T. In this case, $f_+[-c, c] \cap f_-[-c, c] = \emptyset$ (see Piccioni (1990), page 326).

It is easy to see that the derivative of T is given by

$$T'(y) = g_1'(y) = \frac{a(1-b^2)}{(a-by)^2}, \quad \text{if } y \geq 0 \quad \text{and} \quad T'(y) = g_2'(y) = \frac{a(1-b^2)}{(a+by)^2}, \quad \text{if } y < 0. \tag{4.1}$$

¿From (4.1) one observes that g_1 and g_2 are increasing monotone functions.

Let us write $p(x) = \frac{1+bx}{2}$ and $q(x) = 1 - p(x)$. Define the function

$$V(x) = \frac{a(1-b^2)}{2(a-b|x|)}, \quad \text{for } x \in [-c, c]. \tag{4.2}$$

It is easy to see that the function V satisfies the equalities

$$V(f_+(x)) = p(x) = \frac{1+bx}{2} \quad \text{and} \quad V(f_-(x)) = q(x) = \frac{1-bx}{2}.$$

For any $x \in [-c, c]$ define the probability law on $[-c, c]$

$$N(x, dy) = p(x)\delta_{f_+(x)}(dy) + q(x)\delta_{f_-(x)}(dy), \tag{4.3}$$

where δ_z means the unit mass concentrated at z and denote by N the corresponding Markovian kernel on $[-c, c]$. One can rewrite the expression (4.3) in the form

$$N(x, dy) = e^{\log V(f_+(x))}\delta_{f_+(x)}(dy) + e^{\log V(f_-(x))}\delta_{f_-(x)}(dy).$$

By applying the Markovian kernel (4.3) to a function φ one gets the function $\psi = \mathcal{L}(\varphi)$ given by

$$\psi(x) = \mathcal{L}(\varphi)(x) = e^{\log V(f_+(x))}\psi(f_+(x)) + e^{\log V(f_-(x))}\psi(f_-(x)). \tag{4.4}$$

The expression (4.4) is the Ruelle-Perron-Frobenius operator for the potential $\log V(x)$ (see Parry and Pollicott (1991)). Since $p(x) + q(x) = 1$ then $\mathcal{L}(1) = 1$ and from thermodynamic formalism theory there exists $\nu = \nu_{\log V}$, an invariant probability for T,

such that

$$\mathcal{L}^*(\nu) = \nu,$$

where $\mathcal{L}^*$ denotes the dual function of $\mathcal{L}$. The probability ν is the stationary probability law considered by Piccioni (1990). In the notation of Elton and Piccioni (1992) $\nu = \mu$.

¿From $\mathcal{L}(1) = 1$ (see Lopes and Withers (1993)), it follows that the equilibrium measure ν satisfies the following variational problem

$$0 = P(\log V) = \sup_{\theta \in M(T)} \{h(\theta) + \int \log V(x) d\theta(x)\} = h(\nu) + \int \log V(x) d\nu(x), \tag{4.5}$$

where $h(\theta)$ is the entropy of θ and $P(\log V)$ is the topological pressure of $\log V$.

Consider now the potential $-t \log T'(x)$. One also wants to estimate, for each value of t, the pressure

$$s(t) = P(-t \log T^{'}) = \sup_{\theta \in M(T)} \{h(\theta) - t \int \log T'(x) d\theta(x)\}.$$

It is known (Parry and Pollicott (1991) or Lopes (1990)) that $s(\cdot)$ is convex and $s(t_0) = 0 \Leftrightarrow t_0 = HD(S)$, where S is the support's closure of ν (see also Lemma 2 in Piccioni (1990)).

Now, we will explore a special relationship between $\log V$ and $\log T'$.

In order to estimate $HD(S)$ and $s(t)$ we consider the difference

$$\log V(x) - \frac{1}{2} \log T'(x) = \log \frac{a(1-b^2)}{2(a-b|x|)} - \frac{1}{2} \log \frac{a(1-b^2)}{(a-b|x|)^2} = \frac{1}{2} \log a(1-b^2) + \log \frac{1}{2}. \tag{4.6}$$

It follows from relation (4.6) that $\log V$ and $\frac{1}{2} \log T'$ have the homology property (the last expression does not depend on x). Therefore, they determine the same equilibrium states, in other words, $\nu = \nu_{\frac{1}{2} \log T'}$.

¿From the expressions (4.5) and (4.6) one gets

$$\begin{aligned}
0 = P(\log V) &= \sup_{\theta \in M(T)} \{h(\theta) + \int \log \frac{a(1-b^2)}{2(a-b|x|)} d\theta(x)\} \\
&= \sup_{\theta \in M(T)} \left\{ h(\theta) + \frac{1}{2} \int \log \frac{a(1-b^2)}{(a-b|x|)^2} d\theta(x) + \frac{1}{2} \log a(1-b^2) + \log \frac{1}{2} \right\} \\
&= \sup_{\theta \in M(T)} \left\{ h(\theta) + \frac{1}{2} \int \log \frac{a(1-b^2)}{(a-b|x|)^2} d\theta(x) \right\} + \log \frac{\sqrt{a(1-b^2)}}{2} \\
&= s(-\frac{1}{2}) + \log \frac{\sqrt{a(1-b^2)}}{2}.
\end{aligned}$$

Therefore,

$$s(-\frac{1}{2}) = P(\frac{1}{2} \log T') = \sup_{\theta \in M(T)} \left\{ h(\theta) + \frac{1}{2} \int \log \frac{a(1-b^2)}{(a-b|x|)^2} d\theta(x) \right\} = \log \frac{2}{\sqrt{a(1-b^2)}}. \tag{4.7}$$

One observes (see Parry and Pollicott (1991)) that the function $s(t) = P(-t \log T')$ satisfies

$$s(0) = \sup_{\theta \in M(T)} h(\theta) = \textit{topological entropy of } T = \log 2$$

and

$$s(-\frac{1}{2}) = \log \frac{2}{\sqrt{a(1-b^2)}} > \log 2,$$

where the above inequality is due to the fact that $a(1-b^2) < 1$. The linear function

$$Q(t) = t \log a(1-b^2) + \log 2 \tag{4.8}$$

also satisfies $Q(0) = \log 2$ and $Q(-1/2) = \log \frac{2}{\sqrt{a(1-b^2)}}$. Furthermore,

$$Q(t) = t \log a(1-b^2) + \log 2 = 0 \Leftrightarrow t = -\frac{\log 2}{\log a(1-b^2)}.$$

As $s(\cdot)$ is a convex function, the value $HD(S)$ is greater than $k = -\frac{\log 2}{\log a(1-b^2)}$ (see Figure 2).

Therefore, one concludes that an estimation for the Hausdorff dimension of the closure of the support of ν is given by

$$k = -\frac{\log 2}{\log a(1-b^2)} < HD(S) \leq 1.$$

¿From the considerations made in Section 3 about the homology property, the probability ν is equal to $\nu_{\frac{1}{2} \log T'}$. From the convexity of $s(t)$ it is easy to see that α, the exponent scale of $\nu_{\frac{1}{2} \log T'} = \nu$, is smaller than k (see Figure 2), that is

$$\alpha < k. \tag{4.9}$$

This follows from the property of α_t described in Section 3.

The above estimation refers to the case $c \leq b$.

This is the end of the proof of Theorem 3.1.

Now we want to estimate more precisely the exponent scale of ν, specifically for the case $b = c$, corresponding to the invariant measure's support not being a Cantor set.

The case $c > b$ will not be analyzed here, since one can not obtain a map T such that f_+ and f_- are the inverse branches of T. In this case, $f_+[-c, c] \cap f_-[-c, c] \neq \emptyset$ and $f_+[-c, c] \cup f_-[-c, c] = [-c, c]$ (see Piccioni (1990), page 326). Therefore, thermodynamic formalism can not be directly applied.

First one observes that the case $b = c$ corresponds to $a = \frac{1+b^2}{2}$. In fact,

$$b = c \Leftrightarrow b = f_+(b) \Leftrightarrow b = a\frac{2b}{1+b^2} \Leftrightarrow a = \frac{1+b^2}{2}.$$

In this case the upper bound estimate k for α in (4.9) is given by

$$k=-\frac{\log 2}{\log a(1-b^2)} \Leftrightarrow k=-\frac{\log 2}{\log \frac{1-b^4}{2}} \Leftrightarrow k=\frac{1}{1-\frac{\log(1-b^4)}{\log 2}}. \tag{4.10}$$

It follows from the expression (4.10), that for any $b \in (0,1)$, $k<1$.

One concludes that the stationary law ν is singular with respect to the Lebesgue measure, since α the exponent scale of ν is smaller than $k<1$.

For the case $b=c$, this is the answer for the question raised by Piccioni (1990): the measure ν is singular with respect to the Lebesgue measure.

REFERENCES

[1] Billingsley, P., *Ergodic Theory and Information*. John Wiley, New York, 1965.

[2] Elton, J. H. and M. Piccioni, "Iterated function systems arising from recursive estimation problems". *Probability Theory and Related Fields* 91 (1992) 103-114.

[3] Falconer, K., *Fractal Geometry*. John Wiley, New York, 1990.

[4] Lalley, S., "Probabilistic Methods in certain counting problems of ergodic theory". *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, T. Bedford, M. Keane and C. Series (eds.), Oxford Univ. Press (1991) 223-258.

[5] Lopes, A., "Entropy and Large Deviations". *NonLinearity* 3 (1990) 527-546.

[6] Lopes, A., "Entropy, Pressure and Large Deviation". *Cellular Automata, Dynamical Systems and Neural Networks*, E. Goles and S. Martinez (eds.), Kluwer, Massachusetts, (1994) 79-146.

[7] Lopes, A. and W. Withers, "Weight-Balanced Measures and Free-Energy for one-dimensional Dynamics". *Forum Mathematicum* 5 (1993) 161-182.

[8] Lopes, A., "The Dimension Spectrum of the Maximal Measure". *SIAM Journal of Mathematical Analysis* 20 (1989) 1243-1254.

[9] Mañé, R., "On the Hausdorff Dimension of the Invariant Probabilities of Rational Maps". *Lecture Notes in Mathematics*, Springer-Verlag 1331 (1990).

[10] Mañé, R., *Ergodic Theory and Differentiable Dynamics*. Springer-Verlag, New York, 1987.

[11] Parry, W. and M. Pollicott, "Zeta functions and the periodic orbit structure of hyperbolic dynamics". *Asterisque* $187-188$ (1990).

[12] Piccioni, M., "On the Asymptotic Behaviour of the Predictor of a Binary Markov Chain". *Bolletino Unione Mat. Ital.*, Serie A 7 (1990) 319-329.

[13] Piccioni, M. and M. Regoli, "The geometry of attractors for a class of iterated function systems". Preprint, 1992.

[14] Ruelle, D, "The Thermodynamic Formalism for Expanding Maps". *Communications in Mathematical Physics* 125 (1989) 239-262.