Fluctuations of the Empirical Entropies of a Chain of Infinite Order*

Davide Gabrielli †
Universitá Dell'Aquila

Antonio Galves ‡
Universidade de São Paulo

Daniela Guiol §
Universidade de São Paulo

August 25, 2003

Abstract

This paper addresses the question of the fluctuations of the empirical entropy of a chain of infinite order. We assume that the chain takes values on a finite alphabet and loses memory exponentially fast. We consider two possible definitions for the empirical entropy, both based on the empirical distribution of cylinders with length $c \log n$, where n is the size of the sample and c is a suitable constant. The first one is the conditional entropy of the empirical distribution, given a past with length growing logarithmically with the size of the sample. The second one is the rescaled entropy of the empirical distribution of the cylinders of size growing logarithmically with the size of the sample. We prove a central limit theorem for the first one. We also prove that the second one does not have Gaussian fluctuations. This solves a problem formulated in Iosifescu (1965).

*Work partially supported by FAPESP (Projeto Temático *Rhythmic patterns, parameter setting and language change*, grant 98/3382-0) and Projeto Temático *Critical phenomena in evolutive processes and equilibrium systems*, grant 99/11962-9)) and CNPq (Project *Probabilistics Tools for Pattern Identification Applied to Linguistics*, grant 465928/2000-5) and is part of the activities of the Núcleo de Excelência *Critical phenomena in probability and stochastic processes* (grant 66.2177/1996-6)

†Work supported by FAPESP (grant 98/11899-2)

‡Work partially supported by CNPq (grant 301301/79)

§Work supported by a CAPES PhD grant

1 Introduction

This paper considers the following question. Suppose we have a long string of symbols produced by a chain of infinite order. How does the empirical entropy estimated from the sample fluctuate around the true entropy of the chain?

By a chain of infinite order, we mean a stationary stochastic process in which at each step the probability governing the choice of a new symbol depends on the entire past. We will assume that this dependency on the past decreases exponentially fast. We will also assume that the symbols belong to a finite alphabet.

We consider two definitions for the empirical entropy. The first one is the conditional entropy of the empirical distribution, given a past with length growing logarithmically with the size of the sample. The second one is the renormalized entropy of the empirical distribution of cylinders. In this case also the size of the cylinders entering in the definition of the empirical entropy grows logarithmically with the size of the sample.

In his 1965 classical article, Iosifescu considered the second definition, in the simpler case in which the length of the cylinders remains constant and does not grow with the size of the sample. In this case, he proved that a central limit theorem holds. He observed at the end of the paper that the problem remained open when the size of the cylinders is an increasing function of the size of the sample. That is the situation we consider here. We prove that in this case, the empirical entropy defined in the second way does not have Gaussian fluctuations around the theoretical entropy of chain.

In opposition to this negative result, we prove that the central limit theorem holds for the empirical conditional entropy. These two theorems are the main results of this paper.

Chains of infinite order seem to have been first studied by Onicescu and Mihoc (1935) who called them *chains with complete connections* (*chaînes à liaisons complètes*). Their study was soon taken up by Doeblin and Fortet (1937) who proved the first results on speed of convergence toward the invariant measure. The name chains of infinite order was coined by Harris (1955). We refer the reader to Iosifescu and Grigorescu (1990) for a complete survey, and to Fernández, Ferrari and Galves (2001) for an elementary presentation of the subject from a constructive point of view.

Fluctuations of empirical entropies were already studied in Basarin (1959) for sequences of independent random variables and Markov Chains. For chains of infinite order, Iosifescu (1965) proved a central limit theorem for the density of entropy of the empirical k-marginals when k is fixed and does not change with n, the size of the sample. The convergence of the empirical entropy when k increases with n was proved in Ornstein and Weiss (1990) (see also Shields 1996). For a presentation of the problem of the estimation of the entropy from a physical point of view we refer the reader to Schurmann and Grassberger (1996).

To prove our theorems we consider the following strategy. We decompose the difference between the empirical and the theoretical entropies as the sum of the relative entropy between the empirical and theoretical marginals plus a Birkhoff sum of some cylindric functions and a remainder. It turns out that the rate of convergence is different for each one of the empirical entropies we consider. The first one converges as the conditional entropy of the marginals of the source, while the second one converges as the density of entropy of the marginals which is much slower. The basis of both proofs is a Central Limit Theorem for cylinder functions with supports of increasing lengths. This theorem is interesting by itself. Its proof uses a regenerative construction of the chain of infinite order.

The paper is organized as follows. Notation, definitions and the statement of the two main theorems are given in section 2. The proof of the Central Limit Theorem is given in section 3. In section 4 we prove the asymptotic normality of the fluctuations of the empirical conditional entropy. Finally, in section 5 we prove that the density of entropy of the empirical marginals cannot have Gaussian fluctuations except in the case of a sequence of independent and identically distributed random variables.

2 Definitions and main results

Let A be a finite alphabet. Given $x := (x_i)_{i\in\mathbb{Z}} \in A^{\mathbb{Z}}$ and two integers $m \le n$, the finite sequence $(x_m, ..., x_n)$ of elements of A will be denoted x_m^n. We will use also the symbol x_m^n with m or n or both infinite. A sequence $x_{-\infty}^0 := (x_i)_{i\le 0}$ will be called a *history*. Given two histories $x_{-\infty}^0$ and $y_{-\infty}^0$, we say that $x_{-\infty}^0 \overset{m}{=} y_{-\infty}^0$, if $x_i = y_i$ for all $i = 0, -1, \ldots, -m$.

Let $X := (X_i)_{i\in\mathbb{Z}}$ be a stationary process taking values on the finite alphabet A and defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We denote ν the law of the chain. This is the unique measure defined on $A^{\mathbb{Z}}$ such that,

$$\nu([a_1^k]) = \mathbf{P}(X_1^k = a_1^k)\,,$$

for any k and any sequence a_1^k, where $[a_m^n]$ denotes the cylinder set

$$[a_m^n] := \{x \in A^{\mathbb{Z}} : x_m^n = a_m^n\}.$$

Let ν_k be the marginal of size k of ν. It is a probability measure on A^k defined by

$$\nu_k(a_1^k) = \nu([a_1^k])\,.$$

We will also use the shorthand

$$\nu_{k+1}(a_0|a_{-k}^{-1}) = \mathbf{P}(X_0 = a_0|X_{-k}^{-1} = a_{-k}^{-1})\,.$$

The extensive entropy of order k is defined as

$$H(\nu_k) := -\sum_{a_1^k\in A^k} \nu_k(a_1^k)\log\nu_k(a_1^k)\ .$$

The conditional entropy of order k is defined as

$$h(\nu_{k+1}) := -\sum_{a_{-k}^0\in A^{k+1}} \nu_{k+1}(a_{-k}^0)\log\nu_{k+1}(a_0|a_{-k}^{-1})\,,$$

for $k \ge 1$ and as

$$h(\nu_1) := -\sum_{a_0\in A} \nu_1(a_0)\log\nu_1(a_0)\,,$$

for $k = 0$. The entropy of the chain is defined as

$$h(\nu) := \lim_{k\to\infty}\frac{1}{k}H(\nu_k) = \lim_{k\to\infty} h(\nu_k)\,.$$

The fact that the two limits in the definition coincide is a well known result (see for instance Shields 1996). Whenever the reference measure ν is clearly indicated by the context we will use the short notation

$$H_k = H(\nu_k)\,,\ h_k = h(\nu_{k+1})\,,\ h = h(\nu)\,. \tag{2.1}$$

In this paper we are concerned with the estimation of h, given a sample of the chain. Let $x_1^n \in A^n$ be a finite sample of the chain $X = (X_i)_{i\in\mathbb{Z}}$ and take $k \le n$. The empirical k-distribution, given the sample, is the measure on A^k defined as

$$\hat\nu_k(a_1^k; x_1^n) := \frac{1}{n-k+1}\sum_{i=1}^{n-k+1} \mathbf{1}(x_i^{i+k-1} = a_1^k)\,,$$

for all $a_1^k \in A^k$, where $\mathbf{1}(x_i^{i+k-1} = a_1^k)$ denotes the indicator function. We will also use the short notation

$$\hat\nu_{k,n}(\cdot) = \hat\nu_k(\cdot; x_1^n).$$

The extensive empirical entropy of order k is defined as

$$\hat{H}_{k,n} := H(\hat{\nu}_{k,n}) \,,$$

and the conditional empirical entropy of order k is defined as

$$\hat{h}_{k,n} := h(\hat{\nu}_{k+1,n}).$$

We will assume that the process $(X_n)_{n\in\mathbb{Z}}$ satisfies the following conditions

(A) For all $a_1^n \in A^n$

$$\nu_n(a_1^n) > 0.$$

(B) The limit

$$\lim_{m\to\infty} \nu_{m+1}(a_0|a_{-m}^{-1}) := \nu(a_0|a_{-\infty}^{-1})$$

exists for all $a_{-\infty}^0 \in A^{-\mathbb{N}}$.

(C) There is a sequence $(\gamma_m)_{m\in\mathbb{N}}$ with $\gamma_m \searrow 0$ when $m \to +\infty$ such that

$$\sup_{\{a,b \,:\, a_{-\infty}^0 \overset{m}{=} b_{-\infty}^0\}} \left| \frac{\nu(a_0|a_{-\infty}^{-1})}{\nu(b_0|b_{-\infty}^{-1})} - 1 \right| = \gamma_m.$$

We remark that for i.i.d. random variables $\gamma_0 = 0$ and for Markov chains of order $k \geq 1$ condition C is satisfied with $\gamma_m = 0$, for all $m \geq k$. We will call $(X_n)_{n\in\mathbb{Z}}$ a *chain of infinite order* if $\gamma_m > 0$ for any m. Whenever $\gamma_m \leq Me^{-cm}$ where M and c are suitable positive constants we will say that the chain loses memory exponentially fast with rate (at least) c. This will be always the case in this article.

For any measurable function $f : A^{\mathbb{Z}} \to \mathbb{R}$ we will use the shorthand $\mathbf{E}\left[f\right]$ to denote $\mathbf{E}\left[f\left(X_{-\infty}^{+\infty}\right)\right]$. We will also use $\mathbf{E}\left[T_d f\right]$ to denote $\mathbf{E}\left[f\left(T_d X_{-\infty}^{+\infty}\right)\right]$ where T_d denotes the d-step shift $(T_d x)_i = x_{i+d}$, for $d \in \mathbb{Z}$. We define

$$\sigma_f^2(m) := \mathbf{E}\left[\left(\frac{1}{\sqrt{m}}\sum_{d=1}^{m}(T_d f - \mathbf{E}[f])\right)^2\right].$$

A direct computation shows that

$$\sigma_f^2(m) = C_f(0) + 2\sum_{d=1}^{m-1} C_f(d) - \frac{2}{m}\sum_{d=1}^{m-1} dC_f(d), \tag{2.2}$$

where

$$C_f(d) = \mathbf{E}\left[fT_d f\right] - \mathbf{E}\left[f\right]\mathbf{E}\left[f\right].$$

If

$$\sum_{d=1}^{+\infty} dC_f(d) < +\infty \,, \tag{2.3}$$

then

$$\sigma_f^2 := \lim_{m\to\infty} \sigma_f^2(m) = C_f(0) + 2\sum_{d=1}^{+\infty} C_f(d). \tag{2.4}$$

We recall that a chain of infinite order losing memory exponentially fast satisfies condition 2.3 (cf. [BFG]).

The main results of this paper are the following theorems.

Theorem 2.1. *Let $X = (X_i)_{i\in\mathbb{Z}}$ be a stationary process satisfying conditions A, B and C and loosing memory exponentially fast with rate $c > \log|A|$. Let us introduce the function $\phi(x) := -\log\nu(x_0|x_{-\infty}^{-1})$. For any sequence $(k(n))_{n\in\mathbb{N}}$ of positive integers such that $\liminf\frac{k(n)}{\log n} > \frac{1}{2c}$ and $\limsup\frac{k(n)}{\log n} < \frac{1}{2\log|A|}$ the following statements hold.*

- *If $\sigma_\phi^2 > 0$, then*

$$\sqrt{\frac{n}{\sigma_\phi^2}}\left(\hat{h}_{k(n),n} - h\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\,,$$

where $\xrightarrow{\mathcal{D}}$ means convergence in distribution.

- *If $\sigma_\phi^2 = 0$, then*

$$\sqrt{n}\left(\hat{h}_{k(n),n} - h\right) \xrightarrow{\mathbf{P}} 0\,,$$

where $\xrightarrow{\mathbf{P}}$ means convergence in probability.

Theorem 2.2. *Let $X = (X_i)_{i\in\mathbb{Z}}$ be a stationary process satisfying conditions A, B and C and losing memory exponentially fast with $\gamma_0 > 0$ and let $(q(n))_{n\in\mathbb{N}}$ be a sequence of positive real numbers. For any sequence $(k(n))_{n\in\mathbb{N}}$ of positive integers such that $k(n) \to +\infty$ when $n \to +\infty$ and such that $\limsup\frac{k(n)}{\log n} < \frac{1}{2\log|A|}$, the following statements hold.*

- *If $\displaystyle\lim_{n\to\infty}\frac{q(n)}{k(n)} = 0$, then*

$$q(n)\left(\frac{1}{k(n)}\hat{H}_{k(n),n} - h\right) \xrightarrow{\mathbf{P}} 0.$$

- *If $\displaystyle\lim_{n\to\infty}\frac{q(n)}{k(n)} = \alpha,\ 0 < \alpha < +\infty$, then*

$$q(n)\left(\frac{1}{k(n)}\hat{H}_{k(n),n} - h\right) \xrightarrow{\mathbf{P}} \alpha\sum_{i=0}^{\infty}(h_i - h)\,, \tag{2.5}$$

where h and h_i are the entropies defined in 2.1.

- *If $\displaystyle\lim_{n\to\infty}\frac{q(n)}{k(n)} = +\infty$, then for any $r \in \mathbb{R}$*

$$\lim_{n\to\infty}\mathbf{P}\left\{q(n)\left(\frac{1}{k(n)}\hat{H}_{k(n),n} - h\right) < r\right\} = 0.$$

The proofs of Theorems 2.1 and 2.2 are based in the following Central Limit Theorem which is interesting by itself. Before stating this theorem we need some extra notation.

Given a function

$$a_{-\infty}^0 \in A^{-\mathbb{N}} \Rightarrow f(a_{-\infty}^0) \in \mathbb{R}$$

we define its m-variation as

$$v_m(f) := \sup_{\{a,b\,:\,a_{-\infty}^0 \overset{m}{=} b_{-\infty}^0\}} |f(a) - f(b)|\,.$$

Given a function $f : A^{\mathbb{Z}} \to \mathbb{R}$ we define its uniform norm as

$$||f||_\infty = sup_{x\in A^{\mathbb{Z}}}|f(x)|\,.$$

Theorem 2.3. *Let $X = (X_i)_{i\in\mathbb{Z}}$ be a stationary process satisfying conditions A, B and C and losing memory exponentially fast, and such that $\sigma_\phi^2 > 0$, where ϕ is defined as in Theorem 2.1. Let also $(f_k)_{k\in\mathbb{N}}$ be a family of uniformly bounded cylindric functions $f_k(x) = f_k(x_{-k}^0) : A^{k+1} \to \mathbb{R}$, satisfying the conditions*

$$\lim_{k\to+\infty} ||\phi - f_k||_\infty = 0 \tag{2.6}$$

and

$$\sup_{m\in\mathbb{N}} v_k(f_m) \le M\gamma_k, \ , \tag{2.7}$$

for all $k \in \mathbb{N}$, where M is a positive constant.
Then for any sequence $(k(n))_{n\in\mathbb{N}}$ of positive integers such that $k(n) \to +\infty$ when $n \to +\infty$ and such that $\lim_{n\to+\infty} \frac{k(n)^8}{n} = 0$, then we have

$$\frac{1}{\sqrt{\sigma_\phi^2 n}} \sum_{i=1}^{n} \left\{T_i f_{k(n)} - \mathbf{E}\left[f_{k(n)}\right]\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\,.$$

We will use the letter M to denote all the constants appearing in this paper, not only the constant appearing in condition 2.7.
Remarks.
1) In Theorem 2.1, if X is a Markov chain of order m, then the lower bound for $k(n)$ becomes simply $\liminf k(n) \ge m$.
2) Theorem 2.2 remains true also if $k(n) \to K < +\infty$ when $n \to +\infty$, with the only difference that the right hand side of (2.5) is replaced by $\alpha \sum_{i=0}^{K}(h_i - h)$.
3) For the case of i.i.d. processes ($\gamma_0 = 0$), Basarin (1959) proved that the Central Limit Theorem holds for the extensive empirical entropy when $q(n) = (n/\sigma_\phi^2)^{1/2}$.

3 Proof of Theorem 2.3

We start the proof with a lemma on the variances associated with the functions f_k.

Lemma 3.1. *Let $X = (X_i)_{i\in\mathbb{Z}}$ be a stationary process and $(f_k)_{k\in\mathbb{N}}$ be a sequence of cylindric functions and let us assume that both satisfy the hypotheses of Theorem 2.3. Then for any sequence $(k(n))_{n\in\mathbb{N}}$ diverging to $+\infty$ when $n \to \infty$, we have*

$$\lim_{n\to+\infty} \sigma^2_{f_{k(n)}}(n) = \sigma_\phi^2.$$

Proof. Using expressions 2.2 and 2.4 we observe that

$$\begin{aligned} |\sigma^2_{f_{k(n)}}(n) - \sigma_\phi^2| \quad &\le \quad |C_{f_{k(n)}}(0) - C_\phi(0)| + 2\sum_{d=1}^{n-1} |C_{f_{k(n)}}(d) - C_\phi(d)| \\ &\quad +2\sum_{d=n}^{+\infty} |C_\phi(d)| + \frac{2}{n}\sum_{d=1}^{n-1} d|C_{f_{k(n)}}(d)|\,. \end{aligned} \tag{3.1}$$

By direct computation, we have for any $d \ge 0$

$$|C_{f_{k(n)}}(d) - C_\phi(d)| \le M||f_{k(n)} - \phi||_\infty. \tag{3.2}$$

¿From this and the condition 2.6 it follows that

$$\lim_{n\to\infty} |C_{f_{k(n)}}(0) - C_\phi(0)| = 0\,.$$

Let now $(g(n))_{n\in\mathbb{N}}$ be a diverging sequence of positive integers such that

$$\lim_{n\to\infty} g(n)||f_{k(n)}-\phi||_\infty = 0\,. \tag{3.3}$$

The second term in the right hand side of 3.1 can be decomposed as

$$2\sum_{d=1}^{n-1}|C_{f_{k(n)}}(d)-C_\phi(d)| = 2\sum_{d=1}^{g(n)}|C_{f_{k(n)}}(d)-C_\phi(d)| + 2\sum_{d=g(n)+1}^{n-1}|C_{f_{k(n)}}(d)-C_\phi(d)| \tag{3.4}$$

It follows directly from 3.2 and 3.3 that the first term in the right hand side of 3.4 converges to 0 as n diverges.

To obtain an upper bound for the remaining terms in 3.1 and in 3.4 we use the well known fact that under the conditions of Theorem 2.3 there exist two positive constants M and η such that

$$|C_{f_k}(d)| \leq Me^{-\eta d} \quad \text{and} \quad |C_\phi(d)| \leq Me^{-\eta d}\,.$$

This follows for instance from the main Theorem of [BFG]. This concludes the proof of the lemma. ■

The proof of the Central Limit Theorem uses the fact that under the hypotheses of Theorem 2.3 the chain can be constructed with a regenerative scheme. A regenerative scheme is a coupled construction of the chain $X=(X_i)_{i\in\mathbb{Z}}$ together with a stationary renewal process $(\tau_i)_{i\in\mathbb{Z}}$ in such a way that the random strings

$$\{(X_j\,;\tau_i\leq j<\tau_{i+1})\}_{i\in\mathbb{Z}}$$

are independent and, excepted for $i=-1$, identically distributed. Here we adopt the convention that

$$\tau_{-1}\leq 0<\tau_0\,.$$

The variables τ_i are called regeneration times of the chain. The renewal process can be constructed in such a way that the distance between any two successive regeneration times has all its moments finite. For a detailed and self-contained presentation of this type of construction the reader is referred to [FFG], or to the original papers [Ber], [CFF] and [Lal] .

Given a positive integer s, the decimated renewal process $(\tau_i^s)_{i\in\mathbb{Z}}$ is defined as

$$\tau_i^s=\tau_{si}\,,$$

for every $i\in\mathbb{Z}$. In what follows we will take $s=s(n)$ such that

$$\lim_{n\to\infty}\frac{k(n)^2}{s(n)}=0 \text{ and } \lim_{n\to\infty}\frac{s(n)^4}{n}=0\,, \tag{3.5}$$

this is possible for the hypothesis on the sequence $(k(n))_{n\in\mathbb{N}}$. From now on, whenever there is no danger of confusion, we will write s instead of $s(n)$ and k instead of $k(n)$. We define

$$S_i(n) := \sum_{j=\tau_i^s+k}^{\tau_{i+1}^s-1} T_jf_k - \mathbf{E}\left[\sum_{j=\tau_i^s+k}^{\tau_{i+1}^s-1} T_jf_k\right]\,.$$

This definition makes sense when $s(n)\geq k(n)+1$ which is always the case for n big enough. We recall that with the regenerative scheme the random strings

$$\left\{\left(X_j\,;\tau_i^s\leq j<\tau_{i+1}^s\right)\right\}_{i\in\mathbb{N}}$$

are independent and identically distributed. As a consequence the mean zero random variables $(S_i(n))_{i\in\mathbb{N}}$ are independent and identically distributed. Using the Berry-Esseen's inequality we will show that they satisfy the Central Limit Theorem.

We rewrite

$$\sum_{i=1}^{n}(T_i f_k - \mathbf{E}[f_k]) = \sum_{i=0}^{N(n)} S_i(n) + R(n)\,, \tag{3.6}$$

where

$$N(n) := \sum_{j=1}^{+\infty} \mathbf{1}(\tau_j^s \le n)\,,$$

and $R(n)$ is the remainder.

The remainder can be written as

$$R(n) = \sum_{i=1}^{N(n)} R_i(n) + Q(n)\,,$$

where

$$R_i(n) = \sum_{j=\tau_i^s}^{\tau_i^s+k-1} (T_j f_k) - \mathbf{E}\left[\sum_{j=\tau_i^s}^{\tau_i^s+k-1} (T_j f_k)\right]$$

and $Q(n)$ is the remainder of the remainder.

Lemma 3.2. *Under the hypotheses of Theorem 2.3 we have*

$$\frac{R(n)}{\sqrt{n}} \xrightarrow{\mathbf{P}} 0\,.$$

Proof. When n is big enough also the random variables $(R_i(n))_{i\in\mathbb{N}}$ are i.i.d.. Using Wald's identity and Chebychev's inequality we have

$$\mathbf{P}\left(\left|\frac{\sum_{i=1}^{N(n)} R_i(n)}{\sqrt{n}}\right| > \delta\right) \le \frac{\mathbf{E}[N(n)]\,\mathbf{E}\left[(R_i(n))^2\right]}{\delta^2 n}\,. \tag{3.7}$$

To conclude that the right hand side of inequality 3.7 converges to 0 we first observe that by the renewal property

$$\lim_{n\to\infty} \mathbf{E}\left[\frac{N(n)s(n)}{n}\right] = \frac{1}{\mathbf{E}[\tau_1 - \tau_0]} > 0. \tag{3.8}$$

Therefore there exists a positive constant M such that

$$\mathbf{E}[N(n)] \le M\frac{n}{s(n)} \tag{3.9}$$

for all n big enough. Since the functions f_k are uniformly bounded above there exist a positive constant M such that

$$\mathbf{E}\left[(R_i(n))^2\right] \le Mk(n)^2\,. \tag{3.10}$$

Putting together inequalities 3.9 and 3.10 we conclude that the right hand side of inequality 3.7 is bounded above by $M\frac{k(n)^2}{\delta^2 s(n)}$ where M is a suitable positive constant. One of the conditions satisfied by the sequences $k(n)$ and $s(n)$ implies that this upper bound converges to 0 as n diverges.

The proof that

$$\frac{Q(n)}{\sqrt{n}} \xrightarrow{\mathbf{P}} 0\,,$$

is analogous. This concludes the proof of the lemma. ■

To conclude the proof of the Theorem we must show that

$$\frac{\sum_{i=0}^{N(n)} S_i(n)}{\sqrt{\sigma_\phi^2 n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.11}$$

The law of large numbers implies that

$$\frac{N(n)}{[\mathbf{E}\,[N(n)]]} \xrightarrow{\mathbf{P}} 1\,,$$

where $[\mathbf{E}\,[N(n)]]$ denotes the integer part of the expectation $\mathbf{E}\,[N(n)]$. Therefore, by a standard argument (see, for instance Theorem 17.1 from Billingsley's classical treatise [Bil]), 3.11 will follow from

$$\frac{\sum_{i=0}^{[\mathbf{E}[N(n)]]} S_i(n)}{\sqrt{\sigma_\phi^2 n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.12}$$

To prove 3.12, before using the Berry-Esseen's inequality we need to show that the scaling factor appearing in the statement of the Theorem has the right asymptotic behavior. This is the content of the next lemma.

Lemma 3.3. *With the conditions of Theorem 2.3 the following limit holds*

$$lim_{n\to+\infty}\frac{\mathbf{Var}\left[\sum_{i=0}^{[\mathbf{E}[N(n)]]} S_i(n)\right]}{\sigma_\phi^2 n} = 1\,. \tag{3.13}$$

where $\mathbf{Var}\,[\cdot]$ *denotes the variance.*

Proof. Using the fact that $S_i(n)$ are mean zero independent random variables together with 3.8, it is easy to see that expression 3.13 is equivalent to

$$lim_{n\to+\infty}\frac{\mathbf{E}\left[(S_1(n))^2\right]}{s(n)\mathbf{E}\,[\tau_1-\tau_0]} = \sigma_\phi^2\,. \tag{3.14}$$

To prove 3.14, we first recall that Lemma 3.1 assures that

$$\sigma_\phi^2 = \lim_{n\to+\infty} \sigma_{f_{k(n)}}^2(n)\,. \tag{3.15}$$

Using 3.6, we rewrite the expression inside the limit in the right-hand side of 3.15 as

$$l_n\frac{\mathbf{E}\left[(S_1(n))^2\right]}{s(n)\mathbf{E}\,[\tau_1-\tau_0]}\Big(1+\mathcal{R}(n)\Big) + \frac{\mathbf{E}\,[R(n)^2]}{n}\,, \tag{3.16}$$

where

$$\mathcal{R}(n) = \frac{2\mathbf{E}\left[\left(\sum_{i=0}^{N(n)} S_i(n)\right)R(n)\right]}{\mathbf{E}\,[N(n)+1]\,\mathbf{E}\left[(S_1(n))^2\right]}\,, \tag{3.17}$$

and l_n is a sequence converging to 1 when n diverges. We observe that

$$\lim_{n\to+\infty}\frac{\mathbf{E}\,[R(n)^2]}{n} = 0$$

A proof of this fact, ignoring the remainder $Q(n)$, is given in Lemma 3.2. The remainder can be treated in a similar way.

Using the Cauchy-Schwartz inequality and Wald's identity we obtain the upperbound

$$|\mathcal{R}(n)| \leq M \sqrt{\frac{\mathbf{E}\left[(R_1(n))^2\right]}{\mathbf{E}\left[(S_1(n))^2\right]}} \leq \frac{Mk(n)}{\sqrt{\mathbf{E}\left[(S_1(n))^2\right]}}\,, \tag{3.18}$$

where M is a suitable positive constant.

If

$$\lim_{n\to+\infty} \frac{k(n)}{\sqrt{\mathbf{E}\left[(S_1(n))^2\right]}} = 0\,,$$

then the proof would be concluded. Let us assume that this is not the case. Then for an $\epsilon > 0$ there exist infinitely many values of n such that

$$k(n) > \epsilon\sqrt{\mathbf{E}\left[(S_1(n))^2\right]}\,.$$

However using this inequality inside 3.16 would led to a result which is in contradiction with Lemma 3.1. This concludes the proof of the lemma.

■

To conclude the proof of Theorem 2.3 we will show that

$$\lim_{n\to+\infty}\sup_{r\in\mathbb{R}}\left|\mathbf{P}\left\{\frac{\sum_{i=0}^{[\mathbf{E}[N(n)]]} S_i(n)}{\sqrt{\sigma_\phi^2 n}} \leq r\right\} - \Psi(r)\right| = 0\,, \tag{3.19}$$

where

$$\Psi(r) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{r} e^{-u^2/2}du\,.$$

By Lemma 3.3

$$\frac{\sum_{i=0}^{[\mathbf{E}[N(n)]]} S_i(n)}{\sqrt{\sigma_\phi^2 n}} = l_n \frac{\sum_{i=0}^{[\mathbf{E}[N(n)]]} S_i(n)}{\sqrt{\mathbf{E}\left[(S_1(n))^2\right][\mathbf{E}\left[N(n)+1\right]]}},$$

where $l_n \to 1$, when $n \to +\infty$. Therefore the left hand side of expression 3.19 can be rewritten as

$$\sup_{r\in\mathbb{R}}\left|\mathbf{P}\left\{\frac{\sum_{i=0}^{[\mathbf{E}[N(n)]]} S_i(n)}{\sqrt{\mathbf{E}\left[(S_1(n))^2\right][\mathbf{E}\left[N(n)+1\right]]}} \leq r/l_n\right\} - \Psi(r)\right| \tag{3.20}$$

We modify the expression inside the absolute value of 3.20, by adding and subtracting $\psi(r/l_n)$ and then using the triangle's inequality. The second term of the upper bound obtained this way is just

$$\sup_{r\in\mathbb{R}}|\Psi(r) - \Psi(r/l_n)|$$

which goes to 0 as n diverges. Now we finally use the Berry-Esseen Theorem in the remaining term to obtain the upper bound

$$\sup_{r\in\mathbb{R}}\left|\mathbf{P}\left\{\frac{\sum_{i=0}^{[\mathbf{E}[N(n)]]} S_i(n)}{\sqrt{\mathbf{E}\left[(S_1(n))^2\right][\mathbf{E}\left[N(n)+1\right]]}} \leq r/l_n\right\} - \Psi(r/l_n)\right| \leq \frac{M\mathbf{E}\left[|S_1(n)|^3\right]}{\left(\mathbf{E}\left[(S_1(n))^2\right]\right)^{\frac{3}{2}}\sqrt{[\mathbf{E}\left[N(n)+1\right]]}} \tag{3.21}$$

We refer the reader to the classical treatise of Shiryaev [Sh] for a presentation of the Berry-Esseen Theorem.

The fact that the functions $f_{k(n)}$ are uniformly bounded provides an upper bound for the numerator of 3.21

$$\mathbf{E}\left[|S_1(n)|^3\right] \leq M s(n)^3 \,, \tag{3.22}$$

where M is a positive constant. From 3.8 we obtain the lower bound

$$[\mathbf{E}\,[N(n)+1]] \geq M \frac{n}{s(n)} \,, \tag{3.23}$$

where M is another positive constant. Finally as a byproduct of Lemma 3.3 we obtain the lower bound

$$\mathbf{E}\left[(S_1(n))^2\right] \geq M s(n) \,, \tag{3.24}$$

where M is a third positive constant. Putting together the bounds 3.21 3.22, 3.23 and 3.24 we conclude that

$$\sup_{r\in\mathbb{R}} \left| \mathbf{P}\left\{ \frac{\sum_{i=0}^{[\mathbf{E}[N(n)]]} S_i(n)}{\sqrt{\mathbf{E}\left[(S_i(n))^2\right][\mathbf{E}\,[N(n)+1]]}} \leq r/l_n \right\} - \Psi(r/l_n) \right| \leq M \frac{s(n)^2}{\sqrt{n}} \,.$$

By hypothesis this last upper bound goes to 0 as n diverges, and this concludes the proof of the Theorem.

4 Proof of theorem 2.1

We first consider the case $\sigma_\phi^2 > 0$. Let us introduce the cylinder functions $\phi_k(x) = \phi_k(x_{-k}^0)$ defined by $\phi_k(x) := -\log \nu_{k+1}(x_0|x_{-k}^{-1})$ and define

$$D(n) = \sqrt{n}\left(\hat{h}_{k(n),n} - h\right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(T_i \phi_{k(n)}(x) - \mathbf{E}\left[\phi_{k(n)}\right]\right) \,.$$

It is easy to see that the sequence of functions $(\phi_k)_{k\geq 1}$ satisfies the conditions of Theorem 2.3. Therefore if $\sigma_\phi^2 > 0$, it follows that

$$\frac{1}{\sqrt{\sigma_\phi^2 n}} \sum_{i=1}^{n} \left\{T_i \phi_{k(n)} - \mathbf{E}\left[\phi_{k(n)}\right]\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \,.$$

Therefore to prove Theorem 2.1 it is enough to show that

$$D(n) \xrightarrow{P} 0 \,, \tag{4.1}$$

as n diverges. To prove 4.1 we will decompose $D(n)$ in three parts and show that each one of them converges to 0 in probability. The parts are defined as follows

$$D_1(n) = \sqrt{n}\left(h_{k(n)} - h\right) \,,$$

$$D_2(n) = \sqrt{n}\left[H(\hat{\nu}_{k(n),n}|\nu_{k(n)}) - H(\hat{\nu}_{k(n)+1,n}|\nu_{k(n)+1})\right] \,,$$

where

$$H(\hat{\nu}_{k,n}|\nu_k) = \sum_{a_1^k} \hat{\nu}_{k,n}(a_1^k) \log\left[\frac{\hat{\nu}_{k,n}(a_1^k)}{\nu_k(a_1^k)}\right] \,,$$

and

$$D_3(n) = \frac{\sqrt{n}}{n-k(n)} \sum_{i=k(n)+1}^{n} T_i \phi_{k(n)}(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_i \phi_{k(n)}(x) .$$

The expert reader has already identified $H(\hat{\nu}_{k,n}|\nu_k)$ as the relative entropy between the probability measures $\hat{\nu}_{k,n}$ and ν_k on A^k.

Lemma 4.1. *Let X be a stationary chain satisfying conditions A, B and C and losing memory exponentially fast with rate at least c, where c is any fixed positive real number. For any sequence $(k(n))_{n\in\mathbb{N}}$ of positive integers such that*

$$\liminf_{n\to+\infty} \frac{k(n)}{\log n} > \frac{1}{2c} \tag{4.2}$$

we have

$$\lim_{n\to+\infty} D_1(n) = 0 .$$

Proof. By Jensen's inequality

$$h_k - h \geq 0 . \tag{4.3}$$

By definition

$$h_k - h = \int d\nu(a_{-\infty}^0) \left(\log \frac{\nu(a_0|a_{-\infty}^{-1})}{\nu_{k+1}(a_0|a_{-k}^{-1})} \right) . \tag{4.4}$$

By standard calculus

$$\int d\nu(a_{-\infty}^0) \left(\log \frac{\nu(a_0|a_{-\infty}^{-1})}{\nu_{k+1}(a_0|a_{-k}^{-1})} \right) \leq \int d\nu(a_{-\infty}^0) \left(\frac{\nu(a_0|a_{-\infty}^{-1})}{\nu_{k+1}(a_0|a_{-k}^{-1})} - 1 \right) . \tag{4.5}$$

It follows from hypothesis C that

$$\int d\nu(a_{-\infty}^0) \left(\frac{\nu(a_0|a_{-\infty}^{-1})}{\nu_{k+1}(a_0|a_{-k}^{-1})} - 1 \right) \leq \gamma_k . \tag{4.6}$$

By hypothesis

$$\gamma_k \leq M \exp\{-ck\} , \tag{4.7}$$

where M is a strictly positive constant. Using together expressions 4.3, 4.4, 4.5, 4.6 and 4.7 we obtain

$$0 \leq D_1(k(n)) \leq M\sqrt{n} \exp\{-ck(n)\} .$$

The conclusion of the lemma now follows directly from 4.2. ■

Before proving that $D_2(n) \xrightarrow{\mathbf{P}} 0$ we need the following lemma.

Lemma 4.2. *If X is a stationary chain satisfying conditions A, B , C and losing memory exponentially fast then*

$$\sup_{k,n} \sup_{a\in A^{\mathbb{Z}}} \left(\frac{\sigma^2_{\mathbf{1}[a_1^k]}(n)}{\nu([a_1^k])} \right) < +\infty ,$$

where $\mathbf{1}[a_1^k] = \mathbf{1}(x_1^k = a_1^k)$ denotes the indicator function of the cylinder set $[a_1^k]$.

Proof. ¿From 2.2 we obtain:

$$\sigma^2_{\mathbf{1}[a_1^k]}(n) \leq \left|C_{\mathbf{1}[a_1^k]}(0)\right| + 2\sum_{d=1}^{n-1}\left(1+\frac{d}{n}\right)\left|C_{\mathbf{1}[a_1^k]}(d)\right|. \tag{4.8}$$

We need an upper bound for the right hand side of 4.8. The ϕ-mixing property of the chain implies that

$$|C_{\mathbf{1}[a_1^k]}(d)| \leq M\nu([a_1^k])e^{-\eta d}. \tag{4.9}$$

where η and M are suitable positive constants. A constructive proof of this can be performed using the regenerative construction which was introduced in the proof of Theorem 2.3 (cf. [FFG]). To conclude the proof it is enough to use 4.9 in the right hand side of 4.8. ■

Lemma 4.3. *Let X be a stationary chain satisfying conditions A, B and C and losing memory exponentially fast. For any sequence $(k(n))_{n\in\mathbb{N}}$ of positive integers such that*

$$\limsup_{n\to+\infty}\frac{k(n)}{\log n} < \frac{1}{2\log|A|} \tag{4.10}$$

we have

$$D_2(n) \xrightarrow{\mathbf{P}} 0\,.$$

Proof. By definition

$$H(\hat{\nu}_{k,n}|\nu_k) = \sum_{a_1^k}\hat{\nu}_{k,n}(a_1^k)\log\left[\frac{\hat{\nu}_{k,n}(a_1^k)}{\nu_k(a_1^k)}\right]\,.$$

Therefore by Jensen's inequality $0 \leq H(\hat{\nu}_{k,n}|\nu_k)$. This allows us to use Markov's inequality to get

$$\mathbf{P}\left\{H(\hat{\nu}_{k(n),n}|\nu_{k(n)}) \geq \frac{\delta}{\sqrt{n}}\right\} \leq \frac{\sqrt{n}}{\delta}\mathbf{E}\left[H(\hat{\nu}_{k(n),n}|\nu_{k(n)})\right]\ , \tag{4.11}$$

for any $\delta > 0$.

We need an upper bound for the expectation on the right hand side of inequality 4.11. We use again Jensen's inequality to obtain

$$H(\hat{\nu}_{k,n}|\nu_k) \leq \log\left\{\sum_{a_1^k}\left[\frac{\left(\hat{\nu}_{k,n}(a_1^k)\right)^2}{\nu_k(a_1^k)}\right]\right\}$$

We observe that

$$\sum_{a_1^k}\left[\frac{\left(\hat{\nu}_{k,n}(a_1^k)\right)^2}{\nu_k(a_1^k)}\right] = 1 + \sum_{a_1^k}\frac{(\hat{\nu}_{k,n}(a_1^k)-\nu_k(a_1^k))^2}{\nu_k(a_1^k)}\,. \tag{4.12}$$

Therefore from 4.12 and standard calculus it follows that

$$H(\hat{\nu}_{k,n}|\nu_k)\ \leq \sum_{a_1^k}\frac{(\hat{\nu}_{k,n}(a_1^k)-\nu_k(a_1^k))^2}{\nu_k(a_1^k)}\,. \tag{4.13}$$

Using the definition of $\hat{\nu}_{k,n}$, we can rewrite the right hand side of 4.13 as

$$\frac{1}{n-k+1}\sum_{a_1^k}\frac{1}{\nu_k(a_1^k)}\left\{\frac{1}{\sqrt{n-k+1}}\sum_{i=1}^{n-k+1}\left(\mathbf{1}(x_i^{i+k-1}=a_1^k)-\nu_k(a_1^k)\right)\right\}^2\,. \tag{4.14}$$

Therefore from 4.13 and 4.14 it follows that

$$\mathbf{E}\left[H(\hat{\nu}_{k(n),n}|\nu_{k(n)})\right] \leq \frac{1}{n-k(n)+1}\sum_{a_1^{k(n)}}\frac{\sigma^2_{\mathbf{1}[a_1^{k(n)}]}(n-k(n)+1)}{\nu_{k(n)}(a_1^{k(n)})}\,. \tag{4.15}$$

By Lemma 4.2 the right hand side of 4.15 is bounded above by

$$\frac{M|A|^{k(n)}}{n-k(n)+1}\,, \tag{4.16}$$

where M is a positive constant. Putting together inequalities 4.11, 4.15, 4.16 we conclude that

$$\mathbf{P}\left\{H(\hat{\nu}_{k(n),n}|\nu_{k(n)}) \geq \frac{\delta}{\sqrt{n}}\right\} \leq \frac{M|A|^{k(n)}\sqrt{n}}{\delta\,(n-k(n)+1)}\,.$$

By hypothesis 4.10 this last upper bound converges to 0 as n diverges. This concludes the proof of the lemma. ■

Lemma 4.4. *From the hypotheses of Lemma 4.3 it follows that*

$$D_3(n) \xrightarrow{\mathbf{P}} 0\,.$$

Proof. The result follows immediately from the fact that functions ϕ_k, $k \in \mathbb{N}$ are uniformly bounded. ■

This concludes the proof of Theorem 2.1 in the case $\sigma^2_\phi > 0$.

Let us finally consider the case $\sigma^2_\phi = 0$. With our assumptions, the variance vanishes if and only if the chain is an i.i.d. sequence of random variables with

$$\mathbf{P}\{X_n = a\} = \frac{1}{|A|}\,,$$

for any $a \in A$. In this case

$$\phi_k(x) = \phi(x) = \log|A|\,,$$

for any x and all k. This implies that

$$\frac{1}{\sqrt{\sigma^2_\phi n}}\sum_{i=1}^{n}\left\{T_i\phi_{k(n)} - \mathbf{E}\left[\phi_{k(n)}\right]\right\} = 0$$

and the conclusion of the second part of Theorem 2.1 follows trivially.

5 Proof of theorem 2.2

We follow the same approach as for the proof of Theorem 2.1. We define

$$W(n) = q(n)\left[\frac{1}{k(n)}\hat{H}_{k(n),n} - h\right] - \frac{q(n)}{n}\sum_{i=1}^{n}\left\{T_i\bar{\phi}_{k(n)} - \mathbf{E}[\bar{\phi}_{k(n)}]\right\}\,, \tag{5.1}$$

where the functions $\bar{\phi}_k$ are cylinder functions defined by

$$\bar{\phi}_k := \frac{1}{k+1}\sum_{i=0}^{k}\phi_k$$

and $\phi_k(x) = \phi_k(x_{-k}^0)$ are the cylinder functions introduced in the proof of Theorem 2.1 and defined by $\phi_k(x) := -\log \nu_{k+1}(x_0|x_{-k}^{-1})$, when $k \geq 1$, and $\phi_0(x) := -\log \nu_1(x_0)$.

We will decompose $W(n)$ in three parts defined as follows

$$W_1(n) = q(n)\left(\frac{1}{k(n)}H_{k(n)} - h\right),$$

$$W_2(n) = -\frac{q(n)}{k(n)}H(\hat{\nu}_{k(n),n}|\nu_{k(n)}),$$

and

$$W_3(n) = W(n) - W_1(n) - W_2(n).$$

Let us briefly sketch the proof. The functions $\bar{\phi}_k$ satisfy the conditions of Theorem 2.3 and by assumption we have $\sigma_\phi^2 > 0$. Therefore if we take $q(n) = \sqrt{n}$, then the second term of right hand side of 5.1 converges to a normal distribution with variance σ_ϕ^2. At this point we could try to reproduce the proof of Theorem 2.1 and to show that the remainder $W(n)$ vanishes as n diverges. This would imply that the first term of right hand side of 5.1 is asymptotically normal.

The second and the third terms of the remainder both converge to 0 in probability, when $q(n) = \sqrt{n}$. However, with this choice of $q(n)$, the term $W_1(n)$ diverges to $+\infty$.

That's the end of the history. There is no scaling factor which simultaneously assures asymptotic normality for the second term in the right-hand side of 5.1 and convergence to 0 for the remainder term. This impossibility is due to the slow rate of convergence of the extensive entropy, much slower than the rate for the conditional entropy. This is content of the next lemmas. In all of them we will assume that X is a chain satisfying the hypotheses of Theorem 2.2.

Lemma 5.1. *For any sequence $(k(n))_{n\in\mathbb{N}}$ of positive integers such that $k(n) \to +\infty$ when $n \to +\infty$ the following statements hold.*

- *If* $\lim_{n\to\infty} \frac{q(n)}{k(n)} = 0$, *then*

$$\lim_{n\to+\infty} W_1(n) = 0.$$

- *If* $\lim_{n\to\infty} \frac{q(n)}{k(n)} = \alpha,\ 0 < \alpha < +\infty$, *then*

$$\lim_{n\to+\infty} W_1(n) = \alpha \sum_{i=0}^{+\infty}(h_i - h),$$

with $0 < \sum_{i=0}^{+\infty}(h_i - h) < +\infty$.

- *If* $\lim_{n\to\infty} \frac{q(n)}{k(n)} = +\infty$, *then*

$$\lim_{n\to+\infty} W_1(n) = +\infty.$$

Proof. By definition, we have that

$$W_1(n) = \frac{q(n)}{k(n)} \sum_{i=0}^{k(n)-1} (h_i - h).$$

Moreover, Jensens inequality implies that $\sum_{i=0}^{+\infty}(h_i - h) \geq 0$. This is actually a strict inequality, since, by assumption, $\gamma_0 > 0$ and therefore the chain X is not a sequence of independent random variables. The fact that $\sum_{i=0}^{+\infty}(h_i - h) < +\infty$ follows easily from expressions 4.4, 4.5 and 4.6. The lemma follows immediately from these facts. ■

Lemma 5.2. *For any sequence $k(n)$ satisfying the hypotheses of Theorem 2.2 and for any sequence $q(n)$ such that $q(n) \leq Mk(n)\sqrt{n}$, with M a positive constant we have that*

$$W_2(n) \xrightarrow{\mathbf{P}} 0\,,$$

as n diverges.

Proof. The proof is identical to the proof of Lemma 4.3. ■

Lemma 5.3. *There exists a positive constant M such that*

$$W_3(n) \leq M\frac{q(n)k(n)}{n}\,.$$

Proof. The result follows immediately from the fact that the functions $\bar{\phi}_k$, $k \in \mathbb{N}$, are uniformly bounded. ■
Lemmas 5.1. 5.2 and 5.3 together with Theorem 2.3 imply that Theorem 2.2 holds for sequences $q(n)$ bounded above by $M\sqrt{n}$, where M is a positive constant.

Lemma 5.4. *If* $\limsup_{n\to+\infty} q(n)/\sqrt{n} = +\infty$ *and* $\lim_{n\to+\infty} q(n)/k(n) = +\infty$*, then for any $r > 0$ we have*

$$\lim_{n\to+\infty} \mathbf{P}\left\{q(n)\left(\frac{\hat{H}_{k(n),n}}{k(n)} - h\right) > r\right\} = 1\,.$$

Proof. By hypothesis for any positive M, we have $q(n) > M\sqrt{n}$ for infinitely many values of n. Let us define

$$\tilde{q}(n) = \min\{q(n), M\sqrt{n}\}\,.$$

Obviously we have

$$\mathbf{P}\left\{q(n)\left(\frac{\hat{H}_{k(n),n}}{k(n)} - h\right) > r\right\} \geq \mathbf{P}\left\{\tilde{q}(n)\left(\frac{\hat{H}_{k(n),n}}{k(n)} - h\right) > r\right\}\,. \tag{5.2}$$

The part already proved of Theorem 2.2 implies that the right hand-side of inequality 5.2 converges to 1. This concludes the proof of Lemma 5.2 as well as the proof of Theorem 2.2. ■

6 Acknowledgments

We thank J.R. Chazottes, P. Collet, D. Duarte, R. Fernández, N.L. Garcia, S. Popov and P. Shields for many interesting discussions on the subject. The authors thank the ZiF of the University of Bielefeld, the Laboratoire de Recherches Informatiques de l'Université de Paris-Sud and the Dipartimento di Fisica dell'Università di Roma *La Sapienza* for their hospitality during the preparation of this paper.

References

[Bas] Basarin, G. P.: On a Statistical Estimate for the Entropy of a Sequence of Independent Random variables. *Theor. Prob. Appl.* **4**, 333–336. (1959).

[Ber] Berbee, H.: Chains with infinite connections: Uniqueness and Markov representation. *Probab. Th. Rel. Fields*, **76**, 243–253. (1987).

[Bil] Billingsley, P.: *Convergence of probability measures*, First edition. Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York (1968).

[BFG] Bressaud, X., Fernández, R., Galves, A.: Decay of Correlation for non Holderian Dynamics. A coupling Approach. *El. Jour. Prob.*, **4**, N 3, 1–19. (1999).

[CFF] Comets, F., Fernández, R., Ferrari, P.A.: Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12** no. 3, 921–943 (2002).

[DF] Doeblin, W., Fortet, R.: Sur les chaînes à liaisons complètes. *Bull. Soc. Math. France*, **65**, 132–148. (1937).

[FFG] Fernández, R., Ferrari, P.A., Galves, A.: *Coupling, renewal and perfect simulation of chains of infinite order*. Lecture Notes for the Vth Brazilian school of Probability, Ubatuba, August 2001. 92 pp. (2001). available at http://www.ime.br/~pablo/publications.html. file N 78.

[Har] Harris, T.E.: On chains of infinite order. *Pacific J. Math.*, **5**:707–724. (1955).

[Ios] Iosifescu, M.: Sampling Entropy for Random Homogeneous Systems with Complete Connections. *Ann. Math. Stat.* **36**, 1433–1436. (1965).

[IG] Iosifescu, M., Grigorescu, S.: *Dependence with Complete Connections and its Applications*, Cambridge University Press, Cambridge. (1990).

[Lal] Lalley, S.P.: Regeneration representation for one-dimensional Gibbs states. *Ann. Prob.*, **14**, 1262–1271. (1986).

[OM] Onicescu, O., Mihoc, G.: Sur les chaînes de variables statistiques. *Bull Sci., Math.*,**59**, 174–92. (1935).

[OW] Ornstein, D.S., Weiss, B.: How sampling reveals a process. *Ann. Prob.*, **18**, 905–930. (1990).

[S] Shields, P.: *The Ergodic Theory of Discrete Sample Paths.* Graduate Studies in Mathematics **13**. (1996).

[Sh] Shiryaev, A.N.: *Probability.* Graduate Texts in Mathematics **95**, 2^d edition, Springer Verlag. (1989)

[SG] Schurmann, T., Grassberger, P.: Entropy Estimation of Symbol Sequences. *Chaos*, **6**, N 3, 414–427. (1996).

Davide Gabrielli
Dipartimento di Matematica Pura ed Applicata,
Universitá Dell'Aquila,
via Vetoio loc. Coppito
67100 L'Aquila,
Italia
e-mail: `gabriell@univaq.it`

Antonio Galves
Instituto de Matemática e Estatística,
Universidade de São Paulo,
Caixa Postal 66281,
05315-970 São Paulo, SP
Brasil
e-mail: galves@ime.usp.br

Daniela Guiol
Instituto de Matemática e Estatística,
Universidade de São Paulo,
Caixa Postal 66281,
05315-970 São Paulo, SP
Brasil
e-mail: dana.rv@bluewin.ch